

CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX Sequence 530 BP; 172 A; 230 C; 76 G; 52 T; 0 U; 0 Other;
SQ

Query Match 65.4%; Score 17; DB 6; Length 530;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGCAATCTACGAA 26
Db 443 CCACGGCGTCGCGCAATCTACGAA 467

RESULT 29
ABQ14534/C
ID ABQ14534 standard; DNA; 530 BP.
XX
XX AC ABQ14534;
XX
XX 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 1125.
XX
XX Human, cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX FN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX DR WPI; 2002-371829/40.
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX FS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX CC This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX Sequence 530 BP; 52 A; 76 C; 230 G; 172 T; 0 U; 0 Other;
SQ

Query Match 65.4%; Score 17; DB 6; Length 530;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGCAATCTACGAA 26
Db 88 CCACGGCGTCGCGCAATCTACGAA 64

RESULT 30
AAA55205/C
ID AAA55205 standard; DNA; 879 BP.
XX
XX AC AAA55205;
XX
XX DT 30-AUG-2000 (first entry)
XX
XX DE C. symbiosum open reading frame nucleotide sequence SEQ ID NO:37.
XX
XX KW Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
XX characterisation; archae; therapeutic; industrial; laboratory; ds.
XX
XX OS Cenarchaeum symbiosum.
XX
XX FN WO200018909-A2.
XX
XX PD 06-APR-2000.
XX
XX PF 29-SEP-1999; 99WO-US022752.
XX
XX PR 29-SEP-1998; 98US-0102294P.
XX
XX PA (DIVE-) DIVERSA CORP.
XX
XX PI Swanson RV, Feldman RA, Schleper C;
XX
XX DR WPI; 2000-293148/25.
XX
XX DR P-PSDB; AAY90930.
XX
XX PT New nucleic acids and proteins isolated from the non-thermophilic
XX crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
XX physiology of these archae and in therapeutic, industrial or laboratory
XX techniques.
XX
XX FS Claim 7; Page 149-151; 210pp; English.
XX
XX CC AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids and
XX proteins isolated from the non-thermophilic crenarchaeote Cenarchaeum
XX symbiosum. The nucleic acids and proteins identified in the present
XX invention are useful in characterising the physiology of these archae and
XX can be used in therapeutic, industrial or laboratory techniques. AAA55227
XX to AAA55260 represent promoter sequences from Cenarchaeum symbiosum.
XX AAA55261 to AAA55269 represent PCR primers and probes used in examples
XX from the present invention
XX
XX SQ Sequence 879 BP; 142 A; 270 C; 279 G; 188 T; 0 U; 0 Other;

Query Match 65.4%; Score 17; DB 3; Length 879;
Best Local Similarity 80.0%; Pred. NO. 2.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGCGCCGGAATCTACGA 25
Db 434 ACCTCGAGAGCGCCGAGTCTACAA 410

Search completed: June 20, 2004, 10:17:36
Job time : 165.522 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1214.98 Seconds

(without alignments)

639.034 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacggcaggccgaatctacgaa 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| C 1 | 20.8 | 80.0 | 298 | CC930679 | ZMVB3c054 |
| C 2 | 19.6 | 75.4 | 570 | CD667796 | NCE833b87 |
| C 3 | 19.6 | 75.4 | 1025 | 29 AG086420 | Pan trogl |
| C 4 | 19.2 | 73.8 | 648 | 12 B1354861 | B1354861 GM28636.5 |

| | | | | | | |
|------|------|------|------|----|----------|------------|
| 5 | 19.2 | 73.8 | 875 | 28 | BZ567119 | pac2-164 |
| C 6 | 18.6 | 71.5 | 367 | 13 | BY077170 | BY077170 |
| C 7 | 18.6 | 71.5 | 376 | 10 | BE523801 | M41G8STM |
| C 8 | 18.6 | 71.5 | 417 | 12 | BG638782 | LD35909.3 |
| C 9 | 18.6 | 71.5 | 458 | 14 | CF507251 | USDA-PP.1 |
| C 10 | 18.6 | 71.5 | 510 | 9 | AI533801 | SD05764.3 |
| C 11 | 18.6 | 71.5 | 519 | 28 | BH885898 | LB00469A. |
| C 12 | 18.6 | 71.5 | 554 | 13 | BQ623129 | USDA-PP.0 |
| C 13 | 18.6 | 71.5 | 564 | 13 | CA140644 | SCJFR205 |
| C 14 | 18.6 | 71.5 | 610 | 14 | CF507469 | USDA-PP.1 |
| C 15 | 18.6 | 71.5 | 618 | 14 | CF653532 | USDA-PP.0 |
| C 16 | 18.6 | 71.5 | 635 | 14 | CF411287 | USDA-PP.1 |
| C 17 | 18.6 | 71.5 | 653 | 28 | BH885920 | LB00482A. |
| C 18 | 18.6 | 71.5 | 656 | 14 | CF411789 | USDA-PP.1 |
| C 19 | 18.6 | 71.5 | 679 | 13 | BQ623363 | USDA-PP.0 |
| C 20 | 18.6 | 71.5 | 691 | 14 | CF837440 | UCRCS03.0 |
| C 21 | 18.6 | 71.5 | 692 | 14 | CF653361 | USDA-PP.0 |
| C 22 | 18.6 | 71.5 | 704 | 13 | BQ624946 | USDA-PP.0 |
| C 23 | 18.6 | 71.5 | 711 | 14 | CF837439 | USDA-PP.0 |
| C 24 | 18.6 | 71.5 | 769 | 14 | CD574354 | UCRCS03.0 |
| C 25 | 18.6 | 71.5 | 799 | 14 | CD574353 | UCRPT01.0 |
| C 26 | 18.2 | 70.0 | 468 | 10 | BF513871 | UI-H-BW1- |
| C 27 | 18.2 | 70.0 | 486 | 10 | BF515709 | UI-H-BW1- |
| C 28 | 18.2 | 70.0 | 538 | 29 | CF189676 | tigr-gss- |
| C 29 | 18.2 | 70.0 | 703 | 14 | CA182954 | SCR2ST315 |
| C 30 | 18.2 | 70.0 | 735 | 10 | BE469759 | IPdK0210 |
| C 31 | 18.2 | 70.0 | 904 | 12 | B1835362 | 603085347 |
| C 32 | 18.2 | 70.0 | 1024 | 13 | BQ218816 | AGENCOURT |
| C 33 | 18.2 | 70.0 | 1134 | 29 | CNS067M1 | T3 end of |
| C 34 | 18.2 | 69.2 | 139 | 13 | BX618520 | EX618520 |
| C 35 | 18.2 | 69.2 | 709 | 10 | BF274843 | GA_Eb002 |
| C 36 | 18.2 | 69.2 | 953 | 10 | BF300258 | 603032166 |
| C 37 | 18.2 | 69.2 | 1041 | 10 | BG031632 | 602299780 |
| C 38 | 18.2 | 69.2 | 1352 | 12 | B1086420 | 602849871 |
| C 39 | 17.8 | 68.5 | 548 | 12 | BM866574 | mgcs006XM |
| C 40 | 17.8 | 68.5 | 1002 | 29 | CNS0579X | Tetraodon |
| C 41 | 17.8 | 68.5 | 1274 | 13 | BU526190 | AGENCOURT |
| C 42 | 17.6 | 67.7 | 371 | 10 | BE359053 | DG1_24_F0 |
| C 43 | 17.6 | 67.7 | 404 | 14 | CA658448 | wawTC_EK0 |
| C 44 | 17.6 | 67.7 | 406 | 10 | BE425356 | wlm0.PK04 |
| C 45 | 17.6 | 67.7 | 451 | 28 | A2132633 | OSUNB010 |
| C 46 | 17.6 | 67.7 | 472 | 28 | CC409166 | PUHST67TB |
| C 47 | 17.6 | 67.7 | 523 | 12 | B1398345 | 949054810 |
| C 48 | 17.6 | 67.7 | 529 | 12 | BM514048 | KY03Q05.Y |
| C 49 | 17.6 | 67.7 | 543 | 29 | CG045764 | PUNK65TD |
| C 50 | 17.6 | 67.7 | 589 | 13 | BQ801290 | WHE2812_F |
| C 51 | 17.6 | 67.7 | 604 | 10 | BE358044 | DG1_24_G0 |
| C 52 | 17.6 | 67.7 | 616 | 14 | CA955017 | k134f08.Y |
| C 53 | 17.6 | 67.7 | 625 | 13 | CA090342 | SCSGAM210 |
| C 54 | 17.6 | 67.7 | 637 | 29 | CG815480 | SOYAL40TH |
| C 55 | 17.6 | 67.7 | 686 | 9 | AU031566 | AU031566 |
| C 56 | 17.6 | 67.7 | 701 | 28 | BZ453364 | BONDN39TF |
| C 57 | 17.6 | 67.7 | 704 | 28 | BH921123 | odj26b06. |
| C 58 | 17.6 | 67.7 | 747 | 14 | CK215056 | FGAS02700 |
| C 59 | 17.6 | 67.7 | 764 | 28 | BH985934 | cen96g11. |
| C 60 | 17.6 | 67.7 | 772 | 29 | CG305234 | OG0BF60TV |
| C 61 | 17.6 | 67.7 | 797 | 29 | CG978096 | ZUAGM75TH |
| C 62 | 17.6 | 67.7 | 797 | 29 | CG366884 | OGWJP50TV |
| C 63 | 17.6 | 67.7 | 799 | 28 | BH899666 | Ots00733 |
| C 64 | 17.6 | 67.7 | 803 | 14 | CA765092 | AF53-RP1 |
| C 65 | 17.6 | 67.7 | 816 | 28 | BZ719123 | PUDAX43TD |
| C 66 | 17.6 | 67.7 | 828 | 28 | CC336342 | OGRIAL27TV |
| C 67 | 17.6 | 67.7 | 846 | 29 | CG203408 | PUIHC63TD |
| C 68 | 17.6 | 67.7 | 853 | 28 | CC336336 | OGIBB82TH |
| C 69 | 17.6 | 67.7 | 863 | 29 | CG189231 | PUIFD05TD |
| C 70 | 17.6 | 67.7 | 868 | 29 | CG293731 | OGXDI89TV |
| C 71 | 17.6 | 67.7 | 890 | 29 | CC697989 | OGWCF64TV |
| C 72 | 17.6 | 67.7 | 896 | 29 | CG862087 | ZMVB3c027 |
| C 73 | 17.6 | 67.7 | 910 | 28 | CC336341 | OGIBB82TV |
| C 74 | 17.6 | 67.7 | 920 | 28 | CG329165 | OGVCE03TV |
| C 75 | 17.6 | 67.7 | 924 | 28 | AZ683774 | ENTWAB4TF |
| C 76 | 17.6 | 67.7 | 924 | 28 | BZ719119 | PUDAX43TB |

| | | | | | | | | | | | | | |
|-------|------|------|------|----|-----------|---------------------|-------|------|------|------|----|-----------|--------------------|
| C 78 | 17.6 | 67.7 | 985 | 29 | CNS01RER | AL157315 Anopheles | C 151 | 17 | 65.4 | 574 | 14 | CF103480 | CF103480 |
| C 79 | 17.4 | 66.9 | 769 | 10 | BF141737 | BF141737 601787601 | C 152 | 17 | 65.4 | 578 | 14 | CB404010 | OSR015H9 |
| C 80 | 17.4 | 66.9 | 1100 | 12 | B1211982 | B1211982 602936449 | C 153 | 17 | 65.4 | 584 | 14 | CD666771 | ETCSTeg30 |
| C 81 | 17.2 | 66.2 | 190 | 29 | CG907394 | CG907394 ZMWB052 | C 154 | 17 | 65.4 | 588 | 13 | BU571686 | 945184B04 |
| C 82 | 17.2 | 66.2 | 267 | 14 | CA742056 | CA742056 wetic.pk0 | C 155 | 17 | 65.4 | 595 | 14 | CD665628 | ETCSTeg32 |
| C 83 | 17.2 | 66.2 | 325 | 9 | AV836085 | AV836085 AV836085 | C 156 | 17 | 65.4 | 596 | 12 | BM086115 | SAH35F11 |
| C 84 | 17.2 | 66.2 | 371 | 9 | AV918494 | AV918494 AV918494 | C 157 | 17 | 65.4 | 597 | 14 | CA179494 | CA179494 SAVPST105 |
| C 85 | 17.2 | 66.2 | 377 | 12 | CG418303 | CG418303 HVSMEK002 | C 158 | 17 | 65.4 | 612 | 28 | BH210205 | SM1-41H20 |
| C 86 | 17.2 | 66.2 | 390 | 13 | CA009415 | CA009415 HU1A24r | C 159 | 17 | 65.4 | 617 | 10 | B9878192 | 601487583 |
| C 87 | 17.2 | 66.2 | 431 | 9 | AJ424000 | AJ424000 AJ424000 | C 160 | 17 | 65.4 | 623 | 14 | CF863327 | PS25008XK |
| C 88 | 17.2 | 66.2 | 433 | 13 | BQ766842 | BQ766842 EBR008_SQ | C 161 | 17 | 65.4 | 625 | 12 | B1238607 | RE35110.5 |
| C 89 | 17.2 | 66.2 | 460 | 14 | CA631465 | CA631465 wleln.pk0 | C 162 | 17 | 65.4 | 625 | 14 | CA405369 | 7 A. bras |
| C 90 | 17.2 | 66.2 | 496 | 12 | BN0918164 | BN0918164 Esp103_SQ | C 163 | 17 | 65.4 | 626 | 12 | B1238566 | RE35056.5 |
| C 91 | 17.2 | 66.2 | 500 | 9 | AM191490 | AM191490 AU191490 | C 164 | 17 | 65.4 | 627 | 29 | CG602473 | ZMWBBC041 |
| C 92 | 17.2 | 66.2 | 508 | 13 | CA006074 | CA006074 HU05114u | C 165 | 17 | 65.4 | 634 | 12 | B1163007 | RE02295.5 |
| C 93 | 17.2 | 66.2 | 518 | 14 | CB354137 | CB354137 ZF001-P00 | C 166 | 17 | 65.4 | 641 | 14 | CD435659 | EL01N0364 |
| C 94 | 17.2 | 66.2 | 586 | 14 | CB8772243 | CB8772243 HC07C05Y | C 167 | 17 | 65.4 | 642 | 14 | CF611457 | LR CD25CF |
| C 95 | 17.2 | 66.2 | 641 | 14 | CD860447 | CD860447 TE-002H11 | C 168 | 17 | 65.4 | 649 | 9 | AI513469 | LD45752.5 |
| C 96 | 17.2 | 66.2 | 660 | 13 | EX073674 | EX073674 BX073674 | C 169 | 17 | 65.4 | 656 | 10 | BE188342 | FVB 060 L |
| C 97 | 17.2 | 66.2 | 699 | 12 | BJ468153 | BJ468153 BJ468153 | C 170 | 17 | 65.4 | 656 | 12 | BI366362 | RE51881.5 |
| C 98 | 17.2 | 66.2 | 702 | 28 | CC071842 | CC071842 CSU-K331r | C 171 | 17 | 65.4 | 659 | 14 | CA224539 | SCCFL600 |
| C 99 | 17.2 | 66.2 | 724 | 13 | BX073676 | BX073676 BX073676 | C 172 | 17 | 65.4 | 660 | 28 | B2892674 | Hm6_0194 |
| C 100 | 17.2 | 66.2 | 750 | 28 | BH987721 | BH987721 Osi11A06 | C 173 | 17 | 65.4 | 667 | 14 | CF864061 | PS25008XB |
| C 101 | 17.2 | 66.2 | 759 | 29 | CNS038G9 | AL240354 Tetraodon | C 174 | 17 | 65.4 | 671 | 12 | B1228894 | RE35552.5 |
| C 102 | 17.2 | 66.2 | 764 | 14 | CA358950 | CA358950 631799 NC | C 175 | 17 | 65.4 | 693 | 29 | CD819340 | 100006504 |
| C 103 | 17.2 | 66.2 | 841 | 9 | AI069114 | AI069114 mgae0005C | C 176 | 17 | 65.4 | 701 | 14 | CD861967 | AZ01.101M |
| C 104 | 17.2 | 66.2 | 861 | 28 | AQ325329 | AQ325329 mgxb0022L | C 177 | 17 | 65.4 | 715 | 14 | CD826857 | BN25.065F |
| C 105 | 17.2 | 66.2 | 969 | 12 | BG259769 | BG259769 602378029 | C 178 | 17 | 65.4 | 715 | 14 | CF393044 | RTDR3.18 |
| C 106 | 17.2 | 66.2 | 981 | 10 | BE739529 | BE739529 601556465 | C 179 | 17 | 65.4 | 721 | 29 | C5633838 | CGYFC79TV |
| C 107 | 17.2 | 66.2 | 1009 | 29 | CC935282 | CC935282 ZMWBBC054 | C 180 | 17 | 65.4 | 734 | 13 | BQ804860 | WHZ3559 G |
| C 108 | 17.2 | 66.2 | 1027 | 10 | BF216448 | BF216448 601884507 | C 181 | 17 | 65.4 | 747 | 28 | BH513243 | BOZP45TF |
| C 109 | 17.2 | 66.2 | 118 | 10 | BF116130 | BF116130 7H77H11.X | C 182 | 17 | 65.4 | 756 | 9 | AA979545 | LD34162.5 |
| C 110 | 17.2 | 66.2 | 163 | 29 | CG08761 | CG08761 OGWDX76TH | C 183 | 17 | 65.4 | 766 | 9 | AA9350536 | LD30373.5 |
| C 111 | 17.2 | 66.2 | 166 | 28 | AZ821274 | AZ821274 2M0094A03 | C 184 | 17 | 65.4 | 778 | 12 | B1114255 | 602862445 |
| C 112 | 17.2 | 66.2 | 229 | 28 | AQ9662931 | AQ9662931 LERG137TR | C 185 | 17 | 65.4 | 788 | 29 | BX163589 | Danio rer |
| C 113 | 17.2 | 66.2 | 240 | 9 | AI188735 | AI188735 AV188735 | C 186 | 17 | 65.4 | 794 | 28 | B2734543 | OGFBI24TM |
| C 114 | 17.2 | 66.2 | 291 | 12 | BM158980 | BM158980 NXIV 042 | C 187 | 17 | 65.4 | 809 | 29 | CG705111 | OGAW18TH |
| C 115 | 17.2 | 66.2 | 331 | 29 | CG795376 | CG795376 SALK 0803 | C 188 | 17 | 65.4 | 814 | 29 | EX148091 | Danio rer |
| C 116 | 17.2 | 66.2 | 340 | 14 | CF845658 | CF845658 PSHB034XA | C 189 | 17 | 65.4 | 815 | 12 | BG684546 | 6026316312 |
| C 117 | 17.2 | 66.2 | 343 | 12 | BI486572 | BI486572 RE70020.5 | C 190 | 17 | 65.4 | 816 | 28 | B2554434 | pacsl-60 |
| C 118 | 17.2 | 66.2 | 348 | 14 | CB827912 | CB827912 LJNEST80F | C 191 | 17 | 65.4 | 820 | 29 | CG938758 | MBED146TF |
| C 119 | 17.2 | 66.2 | 349 | 14 | CB031405 | CB031405 TGEStzyd2 | C 192 | 17 | 65.4 | 827 | 29 | CG044322 | PUEED14TF |
| C 120 | 17.2 | 66.2 | 360 | 9 | AV192051 | AV192051 AV192051 | C 193 | 17 | 65.4 | 863 | 29 | CG044943 | PUCOS18TB |
| C 121 | 17.2 | 66.2 | 376 | 10 | BF870031 | BF870031 IL3-ET011 | C 194 | 17 | 65.4 | 885 | 28 | CG329736 | CG329736 |
| C 122 | 17.2 | 66.2 | 383 | 12 | BG159027 | BG159027 RHIZ2_41 | C 195 | 17 | 65.4 | 904 | 12 | B1106690 | 602885967 |
| C 123 | 17.2 | 66.2 | 399 | 12 | EG274127 | EG274127 WHE2331_H | C 196 | 17 | 65.4 | 923 | 29 | CG679436 | OGFA271TV |
| C 124 | 17.2 | 66.2 | 399 | 14 | CB384983 | CB384983 OSTF015D1 | C 197 | 17 | 65.4 | 940 | 28 | CG29746 | OGFA271TV |
| C 125 | 17.2 | 66.2 | 399 | 14 | CB403971 | CB403971 OSTF015D1 | C 198 | 17 | 65.4 | 967 | 12 | BG251378 | 602364022 |
| C 126 | 17.2 | 66.2 | 400 | 12 | BG413258 | BG413258 ETESted19 | C 199 | 17 | 65.4 | 969 | 29 | CNS02MJK | AL204185 Tetraodon |
| C 127 | 17.2 | 66.2 | 427 | 9 | AU282923 | AU282923 AU282923 | C 200 | 17 | 65.4 | 1008 | 13 | BQ672812 | AGENCOURT |
| C 128 | 17.2 | 66.2 | 441 | 14 | CA658408 | CA658408 w1m0.pk04 | C 201 | 17 | 65.4 | 1058 | 28 | B2577858 | msb2_5608 |
| C 129 | 17.2 | 66.2 | 447 | 10 | BE4393034 | BE4393034 WHE0562_F | C 202 | 17 | 65.4 | 1122 | 29 | CNS06Q2Y | T3_end of |
| C 130 | 17.2 | 66.2 | 455 | 28 | AZ132296 | AZ132296 OSUNB006 | C 203 | 17 | 65.4 | 1142 | 28 | B2552819 | pacsl-60 |
| C 131 | 17.2 | 66.2 | 464 | 28 | AQ535924 | AQ535924 RFI-11-4 | C 204 | 17 | 65.4 | 1153 | 12 | BMS52430 | AGENCOURT |
| C 132 | 17.2 | 66.2 | 473 | 28 | CA109661 | CA109661 CIT-HSP-2 | C 205 | 17 | 65.4 | 1104 | 10 | BE738725 | 601572974 |
| C 133 | 17.2 | 66.2 | 479 | 14 | CB405915 | CB405915 OSTF059E1 | C 206 | 16.8 | 64.6 | 104 | 13 | B0869595 | U048E04 P |
| C 134 | 17.2 | 66.2 | 492 | 29 | CC820351 | CC820351 100001A22 | C 207 | 16.8 | 64.6 | 385 | 12 | BM147165 | TCAAP1Q10 |
| C 135 | 17.2 | 66.2 | 507 | 28 | AQ879656 | AQ879656 HS_4821_A | C 208 | 16.8 | 64.6 | 439 | 13 | BY447416 | BY447416 |
| C 136 | 17.2 | 66.2 | 507 | 28 | AQ879705 | AQ879705 HS_4821_A | C 209 | 16.8 | 64.6 | 582 | 10 | BF252054 | EST419316 |
| C 137 | 17.2 | 66.2 | 510 | 12 | BG516742 | BG516742 ETESted61 | C 210 | 16.8 | 64.6 | 649 | 12 | BG787388 | SEACMC007 |
| C 138 | 17.2 | 66.2 | 516 | 12 | BG904715 | BG904715 TAlr1134D | C 211 | 16.8 | 64.6 | 663 | 12 | B1669703 | 603293250 |
| C 139 | 17.2 | 66.2 | 517 | 13 | CA001330 | CA001330 HS18106u | C 212 | 16.8 | 64.6 | 993 | 29 | CNS05M64 | AL343669 Tetraodon |
| C 140 | 17.2 | 66.2 | 521 | 29 | AG224590 | AG224590 Lotus cor | C 213 | 16.8 | 64.6 | 1060 | 12 | BM924514 | AGENCOURT |
| C 141 | 17.2 | 66.2 | 534 | 14 | CF753771 | CF753771 EST-77-2- | C 214 | 16.6 | 63.8 | 145 | 10 | BF116847 | UY32a10-Y |
| C 142 | 17.2 | 66.2 | 535 | 9 | AI996716 | AI996716 701668055 | C 215 | 16.6 | 63.8 | 174 | 14 | CF556237 | 1115015H1 |
| C 143 | 17.2 | 66.2 | 540 | 28 | BZ895417 | BZ895417 NARP10.01 | C 216 | 16.6 | 63.8 | 179 | 12 | BG487500 | EMI_65_G1 |
| C 144 | 17.2 | 66.2 | 547 | 28 | BH634335 | BH634335 1008057F0 | C 217 | 16.6 | 63.8 | 234 | 14 | CF865821 | tric003X9 |
| C 145 | 17.2 | 66.2 | 556 | 13 | CA005421 | CA005421 HU100703u | C 218 | 16.6 | 63.8 | 236 | 14 | CA647697 | wleln.pk0 |
| C 146 | 17.2 | 66.2 | 564 | 9 | AI456544 | AI456544 LD36595.5 | C 219 | 16.6 | 63.8 | 243 | 13 | BQ700211 | NKEV102 G |
| C 147 | 17.2 | 66.2 | 565 | 9 | AI770273 | AI770273 SAL3_E02 | C 220 | 16.6 | 63.8 | 274 | 29 | CG712165 | CG712165 |
| C 148 | 17.2 | 66.2 | 568 | 14 | CF882197 | CF882197 tric029x1 | C 221 | 16.6 | 63.8 | 307 | 9 | AU070159 | AU070159 |
| C 149 | 17.2 | 66.2 | 570 | 9 | AI770270 | AI770270 SAL3_D04 | C 222 | 16.6 | 63.8 | 315 | 13 | BQ760612 | EBR003 SQ |
| C 150 | 17.2 | 66.2 | 574 | 12 | BU159696 | BU159696 BU159696 | C 223 | 16.6 | 63.8 | 345 | 14 | CA639072 | wleln.pk0 |


```
224 16.6 63.8 355 9 AA745637
c 225 16.6 63.8 357 14 CA651760
226 16.6 63.8 360 14 CB867769
227 16.6 63.8 361 9 AA166277
c 228 16.6 63.8 377 14 CR881912
229 16.6 63.8 377 14 CF886362
230 16.6 63.8 385 14 CA603922
231 16.6 63.8 392 14 CF556236
232 16.6 63.8 396 12 BJ482339
c 233 16.6 63.8 397 29 CC951609
c 234 16.6 63.8 416 13 BQ660741
c 235 16.6 63.8 421 14 CD937856
c 236 16.6 63.8 424 14 CD882523
c 237 16.6 63.8 427 10 BE591405
c 238 16.6 63.8 434 14 CD057595
239 16.6 63.8 444 9 AU183310
240 16.6 63.8 444 9 AU183310
241 16.6 63.8 445 14 CB866388
242 16.6 63.8 446 10 BE018744
243 16.6 63.8 450 9 AL817225
c 244 16.6 63.8 455 14 CB13437
c 245 16.6 63.8 459 28 BZ334215
c 246 16.6 63.8 463 13 BQ658174
c 247 16.6 63.8 466 28 AZ867494
c 248 16.6 63.8 467 9 AL505858
249 16.6 63.8 480 14 CF643680
250 16.6 63.8 502 14 CA180087
251 16.6 63.8 505 14 CF643588
252 16.6 63.8 512 12 BG909982
253 16.6 63.8 514 14 CF564589
c 254 16.6 63.8 527 29 CNG03HIW
c 255 16.6 63.8 528 9 AV632220
256 16.6 63.8 538 14 CF557920
257 16.6 63.8 539 14 CA642829
258 16.6 63.8 531 14 CB866928
259 16.6 63.8 537 12 BG605539
c 260 16.6 63.8 537 14 CF797231
c 261 16.6 63.8 538 14 CF274843
c 262 16.6 63.8 539 14 CA183146
c 263 16.6 63.8 549 14 CA637222
c 264 16.6 63.8 550 14 CD883343
265 16.6 63.8 551 13 BQ080717
266 16.6 63.8 552 13 BQ080717
267 16.6 63.8 553 14 CB866423
268 16.6 63.8 558 14 CA596282
269 16.6 63.8 559 14 CA700439
270 16.6 63.8 562 14 CA221192
271 16.6 63.8 563 29 CC967554
272 16.6 63.8 565 12 BG632274
273 16.6 63.8 566 13 BQ458375
274 16.6 63.8 568 13 CA112709
275 16.6 63.8 570 14 CK131140
276 16.6 63.8 572 14 CF555545
277 16.6 63.8 577 14 CF797396
c 278 16.6 63.8 577 14 CF942440
279 16.6 63.8 578 28 AQ619676
280 16.6 63.8 579 14 CA239443
281 16.6 63.8 580 12 BG487499
282 16.6 63.8 582 9 AV397650
283 16.6 63.8 585 12 BG606523
c 284 16.6 63.8 587 14 CF560327
c 285 16.6 63.8 587 14 CF565633
286 16.6 63.8 587 14 CF565634
c 287 16.6 63.8 590 29 CG209669
c 288 16.6 63.8 595 13 BU999353
c 289 16.6 63.8 598 14 CA733694
c 290 16.6 63.8 598 14 CF565484
c 291 16.6 63.8 599 29 CG146982
c 292 16.6 63.8 600 28 BZ895451
c 293 16.6 63.8 606 14 CA728823
c 294 16.6 63.8 608 14 CF565483
c 295 16.6 63.8 617 14 CF560045
296 16.6 63.8 622 14 CA606145

AA745637 ny69f10.s
CA651760 wein.pk1
CB867769 HCO1C24w
AA166277 ms4d11.r
CF881912 tr1C085xn
CF886362 tr1C030xa
CA603922 wr1.pk003
CF556236 111501SH1
BJ482339 BJ482339
CC951609 BOHC21TR
BQ660741 HI05D16U
CD937856 OV.108821
CD882523 FL.110C04
BE591405 WHE1659-1
CD057595 HO15J18S
AU183310 AU183310
CB866388 HCO7E20w
CB867017 HCO5F03w
BE018744 bb84d07.Y
AL817225 AL817225
CB13437 OML03717
BZ334215 bx77f05.9
BQ658174 MA05G18u
AZ867494 2M0178520
AL505858 AL505858
CF643680 K11.D09.F
CA180087 SCCST300
CF643588 1115092H0
BG909982 TaLR1112E
CF564589 1115092H0
AL244337 Tetraodon
AV632220 AV632220
CF557920 1115032P0
CA642829 wein.pk0
CB866928 HCO5J17w
BG605539 WHE2332.H
CF797231 NCBST3C56
CF274843 NCBST3434
CA183146 SCMCST316
CA637222 wr1.pk00
CD883343 FL.112P21
CF560047 1115052H0
BQ080717 san37C01.
CB866423 HCO7D01w
CA596282 wpa1C.pk0
CA700439 wkmic.pk0
CA221192 SCBGFLA05
CC967554 BO1FA54PF
BQ632274 GR06015.3
BQ458375 HA05G18r
CA112709 SCEQUB106
CK131140 GH05220.3
CF555545 1115009C0
CF797396 NCBST3C57
CF942440 NCBSTqab9
AQ619676 NS.5185.A
CA239443 SCRFPL503
BG487499 EM1.65.G1
AV397650 AV397650
BG606523 WHE2957.C
CF560327 1115055E0
CF565633 1115101G0
CF565634 1115101G0
CG209669 OGVDU59TH
BU999353 HI14D22r
CA733694 wlp1C.pk0
CF565484 1115100P0
CG146982 PU1E032PB
BZ895451 NARPL1.01
CA728823 wd1C.pk0
CF565483 1115100F0
CF560045 1115052H0
CA606145 wr1.pk006

297 15.6 63.8 624 14 CA187672
c 298 15.6 63.8 624 14 CF563749
299 15.6 63.8 624 14 CF563751
300 15.6 63.8 624 14 CF565485

ALIGNMENTS

CC930679 298 bp DNA linear GSS 11-AUG-2003
ZMMBBC0541G09r ZMMBbc Zea mays subsp. mays genomic clone
ZMMBBC0541G09 3', genomic survey sequence.
CC930679
CC930679.1 GI:33581777
GSS.
Zea mays subsp. mays (maize)
Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 298)
Bharti.A.K., Young.S., Kavchok.S., Keizer.G., Bronzino.A.C.,
Rouzar,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGR (2003b)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 124.
Location/Qualifiers
1..298
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBbc0541G09"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBbc"
/note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"

Query Match 80.0%; Score 20.8; DB 29; Length 298;
Best Local Similarity 91.7%; Pred. No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AACGCGAGGCCCGAATCTACGAA 26
|||||
DB 181 AACGCGAGGCCCGAATCTCGAA 158

RESULT 2
CD667796 570 bp mRNA linear EST 23-JUN-2003
LOCUS NCEST3DB97e08.y1 Nc-1 Tachyzoite cDNA Library Neospora caninum cDNA
5', mRNA sequence.
DEFINITION CD667796
ACCESSION CD667796.1 GI:32162466
VERSION EST
KEYWORDS Neospora caninum
ORGANISM Neospora caninum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Neospora.
1 (bases 1 to 570)
```


XhoI; Sized fractionated cDNAs were directly ligated into pOT2. "

ORIGIN

Query Match 73.8%; Score 19.2; DB 12; Length 648;
Best Local Similarity 84.0%; Pred. No. 2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGCGCGGAATCTAGC 25
||||| ||||| ||||| ||||| |||||
Db 577 AGAAGCGCGAGCGCGGAATCTGCCA 601

RESULT 5

BZ567119 875 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164 6781.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164 6781, genomic survey sequence.

ACCESSION BZ567119

VERSION BZ567119.1 GI:27197852

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 875)

AUTHORS Spencer D.H., Raymond C.K., Smith E.E., Sims E.E., Hastings M.,

Burns J.B., Kaul R. and Olsen M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

JOURNAL J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

source

1..875
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164 6781"
/clone_lib="pacs2-164"
/note="Clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN

Query Match 73.8%; Score 19.2; DB 28; Length 875;
Best Local Similarity 87.5%; Pred. No. 2.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGCGCGGAATCTACG 24
||||| ||||| ||||| ||||| |||||
Db 787 ACAACGGCGAGCGCGGAATCTTCG 810

RESULT 6

BY077170/c 367 bp mRNA linear EST 06-DEC-2002
LOCUS BY077170 RIKEN full-length enriched, pooled tissues, adult spleen,
DEFINITION etc. Mus musculus cDNA clone K630008F15 5', mRNA sequence.

ACCESSION BY077170

VERSION BY077170.1 GI:26178644

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 367)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oeato, N., Saito, R., Sazuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Bakalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Harai, A., Hashizume, M., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12465851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, K.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

Location/Qualifiers

1..367

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="K630008F15"

/clone_lib="RIKEN full-length enriched, pooled tissues,

adult spleen, etc."

FEATURES

source

USDA-FP_122000-752 Immature Ovaries from field-collected Valencia Sweet Orange (Citrus sinensis (L.) Osbeck) Citrus sinensis cDNA clone MVF-55_E04 5', mRNA sequence.

CF507251

CF507251.1 GI:34522435

EST.

KEYWORDS

SOURCE

ORGANISM

Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroids II; Sapindales; Rutaceae; Citrus.

REFERENCE

AUTHORS

Chaparro,J.X., Bausher,M.G., Dang,P., Hunter,W.B., McKenzie,C.L., Niedz,R.P. and Shatters,R.G. Jr.

TITLE

Expressed Sequence Tags from Immature Ovaries of Valencia Sweet Orange (Citrus sinensis (L.) Osbeck)

JOURNAL

Unpublished (2003)

COMMENT

Contact: Chaparro, JX

Horticulture and Breeding Research Unit

USDA, ARS, U. S. Horticultural Research Laboratory

2001, South Rock Road, Fort Pierce, FL 34945, USA

Tel: 772 462 5830

Fax: 772 462 5986

Email: jchaparro@ushrl.ars.usda.gov

Seq primer: T3 Primer

Location/Qualifiers

1..458

/organism="Citrus sinensis"

/mol_type="mRNA"

/cultivar="Valencia"

/db_xref="taxon:2711"

/clone="MVP-55_E04"

/tissue_type="Immature ovaries"

/clone_lib="Immature Ovaries from field-collected Valencia Sweet Orange (Citrus sinensis (L.) Osbeck)"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Standard library construction protocols from XhoI; Stratagene cDNA synthesis kit (cat.# 200401-5) and Uni-ZAP XR vector kit (cat.# 237211) were followed using poly(A) RNA."

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 458;

Best Local Similarity 84.0%; Pred. No. 3.2e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCAGGCCGGAATCTACGAA 26

DB 197 CCACCGCAGGCCGGAATCTACGAA 173

RESULT 10

AI533801/c

LOCUS

SDU5764.3prime SD Drosophila melanogaster Schneider L2 cell culture

DEFINITION

SDU5764.3prime SD Drosophila melanogaster cDNA clone SD05764 3prime, mRNA sequence.

ACCESSION

AI533801

VERSION

AI533801.1 GI:4447936

KEYWORDS

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

Harvey,D., Brkstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.

TITLE

BDGP/HMI Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone probably contains an inverted insert. The resulting Poly-T sequence has been removed.

Plate: 57 row: F column: 4

High quality sequence stop: 403.

Location/Qualifiers

1..510

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="SD05764"

/lab_host="DH5-alpha"

/clone_lib="SD Drosophila melanogaster Schneider L2 cell culture POT2"

/note="Vector: POT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

ORIGIN

Query Match 71.5%; Score 18.6; DB 9; Length 510;

Best Local Similarity 84.0%; Pred. No. 3.3e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCAGGCCGGAATCTACGA 25

DB 191 AGAAGCGCAGCGCGGAATCTCCGA 167

RESULT 11

BH885898/c

LOCUS

LB00469a.d SP6.1 Leishmania major Friedlin BAC Library Leishmania major genomic clone LB00469a, genomic survey sequence.

DEFINITION

BH885898

519 bp DNA linear GSS 07-AUG-2002

ACCESSION

BH885898

VERSION

BH885898.1 GI:22130293

KEYWORDS

GSS.

SOURCE

Leishmania major

ORGANISM

Leishmania major

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

REFERENCE

AUTHORS

Myler,P.J., Vogt,C., Munden,H., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal,G., Nelson,S., Seyler,A., Worthey,E., Stuart,X. and Ragland,M.

TITLE

Leishmania major Friedlin BAC End Sequences

JOURNAL

Unpublished (2002)

COMMENT

Other GSSs: LB00469a.d_T7.1

Contact: Myler PJ

Seattle Biomedical Research Institute

4 Nickerson Street, Seattle, WA 98109-1651, USA

Tel: 206 284-8846

Fax: 206 284-0313

Email: mylerpj@sbri.org

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..519

/organism="Leishmania major"

/mol_type="genomic DNA"

/strain="Friedlin"

/db_xref="taxon:5664"

/clone="LB00469a"

/lab_host="E. coli GeneHogs + TrfA"

/note="Vector: pCG270; Site 1: HindIII; Genomic DNA from Leishmania major Friedlin in agarose blocks was partially digested with HindIII, size selected, and ligated with HindIII-digested pCG270 vector DNA. 10368 clones were picked and arrayed in 384- and 96-well plates. Library

construction and arraying was carried out by ResGen Corporation and clones and filters are available from them"

ORIGIN

Query Match 71.5%; Score 18.6; DB 28; Length 519;
Best Local Similarity 84.0%; Pred. No. 3.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGGAGCGCCGAATCTACGAA 26

Db 378 CACGCGGAGCGCGCAATCAACGAA 354

RESULT 12

BQ623129/c

LOCUS USDA-FP 00220 Ridge pineapple sweet orange entire seedling Citrus
DEFINITION sinensis cDNA clone USDA-FP_00220 5', mRNA sequence.

ACCESSION BQ623129

VERSION BQ623129.1 GI:21650298

KEYWORDS EST.

SOURCE Citrus sinensis

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE

AUTHORS Bausher,M., McKendree,W., Dang,P., Chaparro,J., Shatters,R.,

Hunter,W. and Niedz,R.

Expressed sequence tags isolated from entire sweet orange (C.

sinensis L. Osbeck) seedling

JOURNAL

COMMENT Unpublished (2003)

Contact: Michael Bausher

US Horticultural Research

USDA - ARS

2001 South Rock Rd., Fort Pierce, FL 34945, USA

Tel: (772) 462-5918

Fax: (772) 462-5961

Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Primer.

FEATURES

source

1..554

/organism="Citrus sinensis"

/mol_type="mRNA"

/cultivar="Ridge pineapple"

/db_xref="taxon:2711"

/clone="USDA-FP_00220"

/tissue_type="entire seedling"

/dev_stage="50 days after germination"

/lab_host="X11-Blue"

/clone_lib="Ridge pineapple sweet orange entire seedling"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:

XhoI; A high quality EST with at least 200 contiguous

bases at Trace Tuner score of 20 or better"

ORIGIN

Query Match 71.5%; Score 18.6; DB 13; Length 554;
Best Local Similarity 84.0%; Pred. No. 3.4e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGGAGCGCCGAATCTACGAA 26

Db 147 CCACGCGGAGCGCCGAATCTACGAA 123

RESULT 13

CA140644/c

LOCUS

DEFINITION SCJFRT2053G09.b RP2 Saccharum officinarum cDNA clone SCJFRT2053G09

3', mRNA sequence.

ACCESSION CA140644

VERSION CA140644.1 GI:35033841

KEYWORDS
SOURCE
ORGANISM

Saccharum officinarum
Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.

REFERENCE

AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

TITLE

THE LIBRARIES THAT MADE SUCEST

JOURNAL

COMMENT

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bcccenter.fcav.unesp.br

Plate: 053 row: G column: 09

Seq primer: SP6 Promoter primer.

FEATURES

source

1..564

/organism="Saccharum officinarum"

/mol_type="mRNA"

/db_xref="taxon:4547"

/clone="SCJFRT2053G09"

/lab_host="DH10B"

/clone_lib="RT2"

/note="Organ: Root tips (0.3cm-long) from adult plants;

Vector: pSPort1; Site 1: SalI; Site 2: NotI; An

unidirectional cDNA library generated from [Root

tips (0.3cm-long) from adult plants]. cDNA was prepared

from polyA+ mRNA using SuperScript Plasmid System Kit

(Invitrogen). The double-strand cDNAs were fractionated

in a sepharose CL-2B 40cm-columns and fragments sizing

between 0.8 and 1.5 Kb were directionally cloned into the

vector. Details of each source of RNA and library

construction can be obtained at

http://succest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 71.5%; Score 18.6; DB 13; Length 564;
Best Local Similarity 84.0%; Pred. No. 3.4e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGGAGCGCCGAATCTACGAA 26

Db 482 CACGCGGCGGCTACATCTACGAA 458

RESULT 14

CF507469/c

LOCUS

DEFINITION CF507469 610 bp mRNA linear EST 08-SEP-2003

USDA-FP 122000-970 Immature Ovaries from field-collected Valencia

Sweet Orange (Citrus sinensis (L.) Osbeck) Citrus sinensis cDNA

clone MYP-57_H05 5', mRNA sequence.

CF507469

CF507469.1 GI:34522653

KEYWORDS EST.

SOURCE Citrus sinensis

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 610)

Chaparro,J.X., Bausher,M.G., Dang,P., Hunter,W.B., McKenzie,C.L.,

Niedz,R.P. and Shatters,R.G. Jr.

Expressed Sequence Tags from Immature Ovaries of Valencia Sweet

Orange (Citrus sinensis (L.) Osbeck)

JOURNAL

COMMENT Unpublished (2003)

Contact: Chaparro, JX

Horticulture and Breeding Research Unit
 USDA, ARS, U. S. Horticultural Research Laboratory
 2001, South Rock Road, Fort Pierce, FL 34945, USA
 Tel: 772 462 5830
 Fax: 772 462 5986
 Email: jchaparro@usbrl.ars.usda.gov
 Seq primer: T3 Primer

FEATURES

source
 1. .610
 /organism="Citrus sinensis"
 /mol_type="mRNA"
 /cultivar="Valencia"
 /db_xref="taxon:2711"
 /clone="MVP-57 H05"
 /tissue_type="Immature ovaries"
 /clone_lib="Immature Ovaries from field-collected Valencia
 Sweet Orange (Citrus sinensis (L.) Osbeck)"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; Standard library construction protocols from
 StrataGene cDNA synthesis kit (cat.# 200401-5) and Uni-ZAP
 XR vector kit (cat.# 237211) were followed using poly(A)
 RNA."

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 610;
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 CAACGGCGAGCCCGAATCTACGAA 26
 Db 178 CCACGCCGAGCCCGAATAACGAA 154

RESULT 15

CF653532/c
 LOCUS
 DEFINITION
 USDA-FP_003419 Ridge pineapple sweet orange entire seedling Citrus
 sinensis cDNA clone RSE08E08 5', mRNA sequence.

ACCESSION
 VERSION
 CF653532.1 GI:37509341
 EST

SOURCE
 Citrus sinensis

ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Sapindales; Rutaceae; Citrus.
 1 (bases 1 to 618)
 Bausher, M., Shatters, R., Chaparro, J., Dang, P., Hunter, W. and
 Niedz, R.

REFERENCE
 AUTHORS
 Bausher, M., Shatters, R., Chaparro, J., Dang, P., Hunter, W. and
 Niedz, R.

TITLE

An expressed sequence tag (EST) set from Citrus sinensis L. Osbeck
 whole seedlings and the implications of further perennial source
 investigations

JOURNAL
 COMMENT
 Plant Sci. 165, 415-422 (2003)
 Contact: Michael Bausher
 US Horticultural Research

USDA - ARS
 2001 South Rock Rd., Fort Pierce, FL 34945, USA
 Tel: (772) 462-5918
 Fax: (772) 462-5961

Email: mbausher@usbrl.ars.usda.gov
 Seq primer: T3 Universal

FEATURES

source
 1. .618
 /organism="Citrus sinensis"
 /mol_type="mRNA"
 /cultivar="Ridge Pineapple"
 /db_xref="taxon:2711"
 /clone="RSE08E08"
 /tissue_type="entire seedling"
 /dev_stage="50 Days after germination"
 /lab_host="Xf1-Blue"
 /clone_lib="Ridge pineapple sweet orange entire seedling"
 /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 618;
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 CAACGGCGAGCCCGAATCTACGAA 26
 Db 200 CCACGCCGAGCCCGAATAACGAA 176

RESULT 16

CF417287/c
 LOCUS
 DEFINITION
 USDA-FP_115000-217 Citrus sinensis; Insect-damaged immature
 leaves/stems from field collected Valencia orange Citrus sinensis
 cDNA clone IDFL-003_E10 5', mRNA sequence.

ACCESSION
 VERSION
 CF417287.1 GI:34418414
 EST

KEYWORDS
 SOURCE
 Citrus sinensis

ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE
 AUTHORS
 Shatters, R.G. Jr., Chaparro, J.X., Bausher, M.G., Hunter, W.G. and
 Niedz, R.P.

TITLE
 JOURNAL
 COMMENT
 Expressed Sequence Tags from Insect-Damaged Field-Collected Sweet
 Orange (Citrus Sinensis (L.) Osbeck) Immature Leaves and Stems
 Unpublished (2003)
 Contact: Shatters, Jr. RG
 Subtropical Insects Research Unit
 USDA, ARS, U. S. Horticultural Research Laboratory
 2001, South Rock Road, Fort Pierce, FL 34945, USA
 Tel: 772 462 5912
 Fax: 772 462 5986
 Email: rshatters@usbrl.ars.usda.gov
 Seq primer: T3 Primer

FEATURES

source
 1. .635
 /organism="Citrus sinensis"
 /mol_type="mRNA"
 /cultivar="Valencia"
 /isolate="mixed field population"
 /db_xref="taxon:2711"
 /clone="IDFL-003_E10"
 /tissue_type="Flush leaves and stems"
 /clone_lib="Citrus sinensis: Insect-damaged immature
 leaves/stems from field collected Valencia orange"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; Citrus sinensis (L.) Osbeck; Standard library
 construction protocols from StrataGene cDNA synthesis kit
 (cat.# 200401-5) and Uni-ZAP XR vector kit (cat.# 237211)
 were followed using poly(A) RNA."

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 635;
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 CAACGGCGAGCCCGAATCTACGAA 26
 Db 165 CCACGCCGAGCCCGAATAACGAA 141

RESULT 17

BH85920/c
 LOCUS
 DEFINITION
 BH85920
 LB00482a.d.SP6.1 Leishmania major Friedlin BAC Library Leishmania
 major genomic clone LB00482a, genomic survey sequence.

```

ACCESSION      BH885920
VERSION        BH885920.1  GI:22130315
KEYWORDS       GSS.
SOURCE         Leishmania major
ORGANISM       Leishmania major
               Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
               Leishmania.
REFERENCE      1 (bases 1 to 653)
               Myler,P.J., Vogt,C., Munden,H., Robertson,L., Sisk,E.,
               Fazelinia,G., Aggarwal,G., Nelson,S., Seyler,A., Worthey,E.,
               Stuart,K. and Ragland,M.
               Leishmania major Friedlin BAC End Sequences
TITLE          Unpublished (2002)
JOURNAL        Other GSSs: LB00482a.d.T7.1
COMMENT        Contact: Myler PJ
               Seattle Biomedical Research Institute
               4 Nickerson Street, Seattle, WA 98109-1651, USA
               Tel: 206 284-8845
               Fax: 206 284-0313
               Email: mylerpj@sbri.org
               Seq primer: SP6
               Class: BAC ends.
FEATURES       source
               Location/Qualifiers
               1..653
               /organism="Leishmania major"
               /mol_type="genomic DNA"
               /strain="Friedlin"
               /db_xref="taxon:5664"
               /clone="LB00482a"
               /lab_host="E. coli GeneHogs + TrfA"
               /note="Vector: pCG270; Site 1: HindIII; Genomic DNA from
               Leishmania major Friedlin in agarose blocks was partially
               digested with HindIII, size selected, and ligated with
               HindIII-digested pCG270 vector DNA. 10368 clones were
               picked and arrayed in 384- and 96-well plates. Library
               construction and arraying was carried out by ResGen
               Corporation and clones and filters are available from
               them"
ORIGIN
Query Match      71.5%; Score 18.6; DB 28; Length 653;
Best Local Similarity 84.0%; Pred. NO. 3.5e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGAGGCGCGAATCTACGAA 26
DB 372 CACGCGAGGCGCGAATCTACGAA 348

RESULT 18
LOCUS      CF417889/c
DEFINITION USDA-FP_115000-819 Citrus sinensis: Insect-damaged immature
leaves/stems from field collected Valencia orange Citrus sinensis
cDNA clone IDFL_011_C09 5', mRNA sequence.
ACCESSION  CF417889
VERSION     CF417889.1  GI:34419016
KEYWORDS    EST.
SOURCE      Citrus sinensis
ORGANISM    Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE   1 (bases 1 to 656)
AUTHORS     Shatters, R.G. Jr., Chaparro,J.X., Bausher,M.G., Hunter,W.G. and
Niedz,R.P.
TITLE       Expressed Sequence Tags from Insect-Damaged Field-Collected Sweet
Orange (Citrus Sinensis (L.) Osbeck) Immature Leaves and Stems
JOURNAL     Unpublished (2003)
COMMENT     Contact: Shatters, Jr. RG
             Subtropical Insects Research Unit
             USDA, ARS, U. S. Horticultural Research Laboratory

ACCESSION      BH885920
VERSION        BH885920.1  GI:22130315
KEYWORDS       GSS.
SOURCE         Leishmania major
ORGANISM       Leishmania major
               Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
               Leishmania.
REFERENCE      1 (bases 1 to 653)
               Myler,P.J., Vogt,C., Munden,H., Robertson,L., Sisk,E.,
               Fazelinia,G., Aggarwal,G., Nelson,S., Seyler,A., Worthey,E.,
               Stuart,K. and Ragland,M.
               Leishmania major Friedlin BAC End Sequences
TITLE          Unpublished (2002)
JOURNAL        Other GSSs: LB00482a.d.T7.1
COMMENT        Contact: Myler PJ
               Seattle Biomedical Research Institute
               4 Nickerson Street, Seattle, WA 98109-1651, USA
               Tel: 206 284-8845
               Fax: 206 284-0313
               Email: mylerpj@sbri.org
               Seq primer: SP6
               Class: BAC ends.
FEATURES       source
               Location/Qualifiers
               1..653
               /organism="Leishmania major"
               /mol_type="genomic DNA"
               /strain="Friedlin"
               /db_xref="taxon:5664"
               /clone="LB00482a"
               /lab_host="E. coli GeneHogs + TrfA"
               /note="Vector: pCG270; Site 1: HindIII; Genomic DNA from
               Leishmania major Friedlin in agarose blocks was partially
               digested with HindIII, size selected, and ligated with
               HindIII-digested pCG270 vector DNA. 10368 clones were
               picked and arrayed in 384- and 96-well plates. Library
               construction and arraying was carried out by ResGen
               Corporation and clones and filters are available from
               them"
ORIGIN
Query Match      71.5%; Score 18.6; DB 28; Length 653;
Best Local Similarity 84.0%; Pred. NO. 3.5e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGAGGCGCGAATCTACGAA 26
DB 372 CACGCGAGGCGCGAATCTACGAA 348

RESULT 19
LOCUS      BQ623363/c
DEFINITION USDA-FP_00454 Ridge pineapple sweet orange entire seedling Citrus
sinensis cDNA clone USDA-FP_00454 5', mRNA sequence.
ACCESSION  BQ623363
VERSION     BQ623363.1  GI:21650532
KEYWORDS    EST.
SOURCE      Citrus sinensis
ORGANISM    Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE   1 (bases 1 to 679)
AUTHORS     Bausher,M., McKendree,W., Dang,P., Chaparro,J., Shatters,R.,
Hunter,W. and Niedz,R.
TITLE       Expressed sequence tags isolated from entire sweet orange (C.
sinensis L. Osbeck) seedling
JOURNAL     Unpublished (2003)
COMMENT     Contact: Michael Bausher
             US Horticultural Research
             USDA - ARS
             2001 South Rock Rd., Fort Pierce, FL 34945, USA
             Tel: (772) 462-5918
             Fax: (772) 462-5961
             Email: mbausher@ushrl.ars.usda.gov
             Seq primer: T3 Primer.
             Location/Qualifiers
             1..679
             /organism="Citrus sinensis"
             /mol_type="mRNA"
             /cultivar="Ridge Pineapple"
             /db_xref="taxon:2711"
             /clone="USDA-FP_00454"
             /tissue_type="entire seedling"
             /dev_stage="50 days after germination"
             /lab_host="X11-Blue"
             /clone_lib="Ridge pineapple sweet orange entire seedling"
             /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site_2:
             XhoI; A high quality EST with at least 200 contiguous
             bases at Trace Tuner score of 20 or better"

```


ORIGIN

Query Match 71.5%; Score 18.6; DB 13; Length 679;
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCGCAATCTACGAA 26
 Db 198 CCACGCCGAGCCGCAAAATACGAA 174

RESULT 20
CF837440/c

LOCUS CF837440 691 bp mRNA linear EST 30-OCT-2003
 DEFINITION UCRS03_04K20_r Washington Navel Orange Shoot Meristem cDNA Library
 Cactus Sinensis cDNA clone CS_PEA04K20, mRNA sequence.

ACCESSION

VERSION CF837440

KEYWORDS

SOURCE EST.

ORGANISM

Citrus sinensis
 Citrus sinensis
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE

1 (bases 1 to 691)

AUTHORS

Close,T.J., Roose,M.L., Federici,C.F., Mu,L., Fenton,R.D.,
 Wanamaker,S., Kim,H.R., Kudrna,D., Wing,R. and Yu,Y.

TITLE

Development of EST Resources and New Genetic Markers for California
 Citrus - Washington Navel Orange Shoot Meristem

JOURNAL

Unpublished (2003)

COMMENT

Contact: Timothy Close

Department of Botany & Plant Sciences, University of California

Riverside, CA, 92521-0124

Tel: 9097873318

Fax: 9097874437

Email: timothy.close@ucr.edu

Seq primer: T3.

FEATURES

Location/Qualifiers

1..691
 /organism="Citrus sinensis"
 /mol_type="mRNA"
 /cultivar="Parent Washington Navel"
 /db_xref="taxon:2711"
 /clone="CS_PEA04K20"
 /tissue_type="Shoot meristem"
 /dev_stage="10 year old trees"
 /lab_host="E. coli TJC121"
 /clone_lib="Washington Navel Orange Shoot Meristem cDNA
 Library"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; Parent Washington Navel
 Orange trees on Troyer rootstock (UCR 16K) were the source
 of tissue. Trees at UC Riverside Agricultural Operations,
 were planted October 12, 1992. In each of 17 reps one tree
 on Troyer rootstock was initially treated with Enzone, one
 with Alliette and Nemacure, and one was left untreated.
 These treatments were discontinued in 1998. At the time of
 sampling, there were differences in the apparent health
 and size of the trees on Troyer rootstock. Fall-flush
 shoots were sampled in early November 2002 to minimize the
 number of floral shoot meristems. Federici and Mu (Roose
 lab) harvested meristems only from trees that appeared to
 be healthy and had a large number of young shoot tips on
 the day of collection. The average weight of a meristem
 was about 2 mg. Federici noted that there were quite a few
 insects and signs of insect damage to the shoot tips.
 Mealy bugs, thrips and aphids were observed, plus a few
 very tiny fast moving insects that may have been mites or
 crawler stage of scale (although Federici did not see any
 mature scale). It was not difficult to avoid collecting
 most of these because they were easy to see with the
 dissecting microscope. It was harder to exclude the frass.
 Some frass was definitely retained in the samples. Tissues

were snap frozen and then stored at -80C until further
 processing. Fenton (Close lab) purified RNA by the phenol
 method described in J. Japanese Soc. Hort. Sci. 1996. 64
 (4): 809-814, purified poly(A) mRNA using a PolyAtrack
 mRNA Isolation System IV (Promega), produced a primary
 cDNA library using a lambda ZAP XR cDNA Synthesis Kit
 (Stratagene), then mass-excised one million pfu from the
 primary library to produce a phagemid population.
 Phagemids were plated, plasmid DNA purified, cDNA clones
 archived, and DNA sequences determined bi-directionally
 using an ABI3730 at the Arizona Genomics Institute,
 University of Arizona (Kim, Kudrna, Wing, Yu).
 Chromatogram files were downloaded by FTP to UC Riverside
 (by Close), then processed at UC Riverside (by Wanamaker,
 Close lab) using the HarVest pipeline
 (http://harvest.ucr.edu) to remove vector and cloning
 oligo sequences and various contaminants, and to trim to a
 high quality region. Sequences that retained a phred 17
 region of at least 100 bases were deposited to GenBank."

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 691;
 Best Local Similarity 84.0%; Pred. No. 3.6e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCGCAATCTACGAA 26

Db 209 CCACGCCGAGCCGCAAAATACGAA 185

RESULT 21

CF653361/c

LOCUS

DEFINITION

USDA-FP_003243 Ridge pineapple sweet orange entire seedling Citrus

sinensis cDNA clone RSE05C05 5', mRNA sequence.

ACCESSION

CF653361

VERSION

CF653361.1

KEYWORDS

SOURCE

ORGANISM

Citrus sinensis

Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE

1 (bases 1 to 692)

AUTHORS

Bausher,M., Shatters,R., Chaparro,J., Dang,P., Hunter,W. and

Niedz,R.

TITLE

An expressed sequence tag (EST) set from Citrus sinensis L. Osbeck

whole seedlings and the implications of further perennial source

investigations

Plant Sci. 165, 415-422 (2003)

CONTACT: Michael Bausher

US Horticultural Research

USDA - ARS

2001 South Rock Rd., Fort Pierce, FL 34945, USA

Tel: (772) 462-5918

Fax: (772) 462-5961

Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Universal.

Location/Qualifiers

1..692

/organism="Citrus sinensis"

/mol_type="mRNA"

/cultivar="Ridge pineapple"

/db_xref="taxon:2711"

/clone="RSE05C05"

/tissue_type="entire seedling"

/dev_stage="50 days after germination"

/lab_host="Xli-Blue"

/clone_lib="Ridge pineapple sweet orange entire seedling"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:

XhoI; A high quality EST with at least 200 contiguous

bases at Trace Tunes score of 20 or better"

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 692;

Best Local Similarity 84.0%; Pred. No. 3.6e-03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

Db 199 CCACGCCGAGCGCCGAAATACGAA 175

RESULT 22

LOCUS BQ624946/c

DEFINITION

USDA-FP 02037 Ridge pineapple sweet orange entire seedling Citrus sinensis cDNA clone USDA-FP_02037 5', mRNA sequence.

ACCESSION BQ624946

VERSION BQ624946.1

KEYWORDS GI:21652115

SOURCE Citrus sinensis

ORGANISM Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 704)

REFERENCE Bausher, M., McKendree, W., Dang, P., Chaparro, J., Shatters, R.,

Hunter, W. and Niedz, R.,

Expressed sequence tags isolated from entire sweet orange (C.

sinensis L. Osbeck) seedling

Unpublished (2003)

CONTACT: Michael Bausher

US Horticultural Research

USDA - ARS

2001 South Rock Rd., Fort Pierce, FL 34945, USA

Tel: (772) 462-5918

Fax: (772) 462-5961

Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Primer.

FEATURES

Source

1..704

/organism="Citrus sinensis"

/mol_type="mRNA"

/cultivar="Ridge Pineapple"

/db_xref="taxon:2711"

/clone="USDA-FP 02037"

/tissue_type="entire seedling"

/dev_stage="50 days after germination"

/lab_host="XLI-Blue"

/clone_lib="Ridge pineapple sweet orange entire seedling"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:

XhoI; A high quality EST with at least 200 contiguous

bases at Trace Tuner score of 20 or better"

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

Db 194 CCACGCCGAGCGCCGAAATACGAA 170

RESULT 23

LOCUS CF837439

DEFINITION

UCRCS03_04K20.f Washington Navel Orange Shoot Meristem cDNA Library

Citrus sinensis cDNA clone CS_PEA04K20, mRNA sequence.

ACCESSION CF837439

VERSION CF837439.1

KEYWORDS GI:38053091

SOURCE Citrus sinensis

ORGANISM Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 711)

REFERENCE Close, T.J., Roose, M.L., Federici, C.F., Mu, L., Fenton, R.D.,

Wanamaker, S., Kim, H.R., Kudrna, D., Wing, R. and Yu, Y.

Development of EST Resources and New Genetic Markers for California

Citrus - Washington Navel Orange Shoot Meristem

Unpublished (2003)

CONTACT: Timothy Close

Department of Botany & Plant Sciences, University of California

Riverside, CA, 92521-0124

Tel: 9097873318

Fax: 9097874437

Email: timothy.close@ucr.edu

Seq primer: T7.

FEATURES

Source

1..711

/organism="Citrus sinensis"

/mol_type="mRNA"

/cultivar="Parent Washington Navel"

/db_xref="taxon:2711"

/clone="CS_PEA04K20"

/tissue_type="Shoot meristem"

/dev_stage="10 year old trees"

/lab_host="E. coli TJC121"

/clone_lib="Washington Navel Orange Shoot Meristem cDNA

Library"

/note="Vector: Lambda Uni-ZAP XR, excised phagenid;

Site_1: EcoRI; Site_2: XhoI; Parent Washington Navel

Orange trees on Troyer rootstock (UCR 16K) were the source

of tissue. Trees, at UC Riverside Agricultural Operations,

were planted October 12, 1992. In each of 17 reps one tree

on Troyer rootstock was initially treated with Enzone, one

with Alliette and Nemacur, and one was left untreated.

These treatments were discontinued in 1998. At the time of

sampling, there were differences in the apparent health

and size of the trees on Troyer rootstock. Fall-flush

shoots were sampled in early November 2002 to minimize the

number of floral shoot meristems. Federici and Mu (Roose

lab) harvested meristems only from trees that appeared to

be healthy and had a large number of young shoot tips on

the day of collection. The average weight of a meristem

was about 2 mg. Federici noted that there were quite a few

insects and signs of insect damage to the shoot tips.

Mealy bugs, thrips and aphids were observed, plus a few

very tiny fast moving insects that may have been mites or

crawler stage of scale (although Federici did not see any

mature scale). It was not difficult to avoid collecting

most of these because they were easy to see with the

dissecting microscope. It was harder to exclude the frass.

Some frass was definitely retained in the samples. Tissues

were snap frozen and then stored at -80C until further

processing. Fenton (Close lab) purified RNA by the phenol

method described in J. Japanese Soc. Hort. Sci. 1996. 64

(4): 809-814. purified poly(A) mRNA using a PolyAtrack

mRNA Isolation System IV (Promega), produced a primary

cDNA library using a lambda ZAP XR cDNA Synthesis kit

(Stratagene), then mass-excised one million pfu from the

primary library to produce a phagemid population.

Phagemids were plated, plasmid DNA purified, cDNA clones

archived, and DNA sequences determined bi-directionally

using an ABI3730 at the Arizona Genomics Institute,

University of Arizona (Kim, Kudrna, Wing, Yu).

Chromatogram files were downloaded by FTP to UC Riverside

(by Close), then processed at UC Riverside (by Wanamaker,

Close lab) using the HarVest pipeline

(http://harvest.ucr.edu) to remove vector and cloning

oligo sequences and various contaminants, and to trim to a

high quality region. Sequences that retained a phred 17

region of at least 100 bases were deposited to GenBank."

ORIGIN

Query Match

71.5%; Score 18.6; DB 14; Length 711;

| | | | | | | | | | |
|-----------------------|---|---------------------------|------------|------------|-----------------|--------|------|------|----|
| Best Local Similarity | 84.0%; | Pred. No. 3.6e+03; | Mismatches | 0; | Indels | 4; | Gaps | 0; | |
| Matches | 21; | Conservative | 0; | Mismatches | 4; | Indels | 0; | Gaps | 0; |
| QY | 2 | CAACGCCGAGGCCCGAATCTACGAA | 26 | | | | | | |
| Db | 629 | CCACGCCGAGGCCCGAATCTACGAA | 653 | | | | | | |
| RESULT 24 | | | | | | | | | |
| CD574354 | | 769 bp | mRNA | linear | EST 12-JUN-2003 | | | | |
| LOCUS | CD574354 | | | | | | | | |
| DEFINITION | UCRPT01_05_E12_TV Poncirus trifoliata CTV-challenged cDNA library - | | | | | | | | |
| ACCESSION | UCR Poncirus trifoliata cDNA clone UCRPT01_05_E12, mRNA sequence. | | | | | | | | |
| VERSION | CD574354.1 | GI:31670256 | | | | | | | |
| KEYWORDS | EST. | | | | | | | | |
| SOURCE | Poncirus trifoliata | | | | | | | | |
| ORGANISM | Poncirus trifoliata | | | | | | | | |
| REFERENCE | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | | | | | | | |
| AUTHORS | Wanamaker,S., Choi,Y. and Kingan,T. | | | | | | | | |
| TITLE | Development of EST Resources and New Genetic Markers for California | | | | | | | | |
| JOURNAL | Citrus - Poncirus trifoliata CTV-challenged phloem - UCR | | | | | | | | |
| COMMENT | Unpublished (2003) | | | | | | | | |
| | Contact: Mikeal Roose | | | | | | | | |
| | Department of Botany & Plant Sciences, University of California | | | | | | | | |
| | Riverside, CA, 92521-0124, USA | | | | | | | | |
| | Tel: 9097874137 | | | | | | | | |
| | Fax: 9097874437 | | | | | | | | |
| | Email: mikeal.roose@ucr.edu | | | | | | | | |
| | Seq primer: (T)20V. | | | | | | | | |
| FEATURES | Location/Qualifiers | | | | | | | | |
| source | 1..769 | | | | | | | | |
| | /organism="Poncirus trifoliata" | | | | | | | | |
| | /mol_type="mRNA" | | | | | | | | |
| | /cultivar="Pomeroy Op" | | | | | | | | |
| | /db_xref="taxon:37690" | | | | | | | | |
| | /clone="UCRPT01_05_E12" | | | | | | | | |
| | /tissue_type="Phloem" | | | | | | | | |
| | /dev_stage="10 - 30 cm shoots" | | | | | | | | |
| | /lab_host="E. coli TUC121" | | | | | | | | |
| | /clone_lib="Poncirus trifoliata CTV-challenged cDNA | | | | | | | | |
| | library - UCR" | | | | | | | | |
| | /note="Vector: Lambda Uni-ZAP XR, excised phagemid; | | | | | | | | |
| | Site 1: EcoRI; Site 2: XhoI; plants were grown in the | | | | | | | | |
| | greenhouse at University of California, Riverside. The | | | | | | | | |
| | scion was a open-pollinated (very probably selfed) | | | | | | | | |
| | seedling of Poncirus trifoliata cv Pomeroy that was | | | | | | | | |
| | selected as homozygous for the Ctv resistance gene. The | | | | | | | | |
| | rootstock was sweet orange infected with citrus tristeza | | | | | | | | |
| | virus (CTV) isolate T514 over 1 year before sampling (CTV | | | | | | | | |
| | infests sweet orange, but not genotypes carrying the Ctv | | | | | | | | |
| | resistance gene. Shoots 10-30 cm long were harvested in | | | | | | | | |
| | October 2000, and the green phloem (bark) was removed and | | | | | | | | |
| | frozen quickly in dry ice. Total RNA was extracted using | | | | | | | | |
| | Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA | | | | | | | | |
| | library was made, and 0.5 million primary lambda cDNA | | | | | | | | |
| | clones were in vivo excised to give a population of | | | | | | | | |
| | pBluescript SK(-) phagemids. All steps to this point were | | | | | | | | |
| | performed in the ML Roose lab at the University of | | | | | | | | |
| | California, Riverside by X. Ye. Phagemids were plated, | | | | | | | | |
| | plasmid DNA purified, cDNA clones archived, and DNA | | | | | | | | |
| | sequences determined bi-directionally using an ABI3730 at | | | | | | | | |
| | the University of California Riverside Genomics Institute, | | | | | | | | |
| | Core Instrumentation Facility. (Choi, Kingan). | | | | | | | | |
| | Chromatogram files were transmitted to UC Riverside (by | | | | | | | | |
| | Choi), then processed at UC Riverside (by Wanamaker) using | | | | | | | | |
| | the HarVEST pipeline (http://harvest.ucr.edu) to remove | | | | | | | | |
| | vector and cloning oligo sequences and various | | | | | | | | |
| | contaminants, and to trim to a high quality region. | | | | | | | | |

Chromatogram files were transmitted to UC Riverside (by Choi), then processed at UC Riverside (by Wananaker) using the HarVest pipeline (<http://harvest.ucr.edu>) using vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 799;
Best Local Similarity 84.0%; Pred. No. 3.7e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGGAATCTACGAA 26
DB 184 CCACGCCGAGGCCGGAATCTACGAA 160

RESULT 26
BF513871/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF513871
UI-H-BM1-amq-b-02-0-UI.s1 NCI CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3070778 3', mRNA sequence.

BF513871
BF513871.1 GI:11599050
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 468)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Oligo-dt track not found. Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 18-43,
>GC-rich#Low complexity
Seq primer: M13 Forward
POLYA=No.

FEATURES

source

Location/Qualifiers
1..468
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3070778"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP_Sub7"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub7
is a subcloned library derived from NCI CGAP Sub6. The
NCI CGAP Sub7 library had 12 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI CGAP_Lus pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clones 1414920-1417991,
1520904-1522439); NCI CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI CGAP_Pt22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clones 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP_Col0 pool 1 LLAM 2644-2653,

ORIGIN

Query Match 70.0%; Score 18.2; DB 10; Length 468;
Best Local Similarity 87.0%; Pred. No. 4.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCGGAATCTAC 23
DB 460 ACAACGGAGAGTCCAGATCTAC 438

RESULT 27

BF515709/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF515709
UI-H-BM1-ano-b-02-0-UI.s1 NCI CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3082707 3', mRNA sequence.

BF515709
BF515709.1 GI:11600888
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 486)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Oligo-dt track not found. Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 18-43,
>GC-rich#Low complexity
Seq primer: M13 Forward
POLYA=No.

FEATURES

source

Location/Qualifiers
1..486
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3082707"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP_Sub7"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub7
is a subcloned library derived from NCI CGAP Sub6. The
NCI CGAP Sub7 library had 12 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778

(IMAGE Clonoids 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI CGAP LAM5 pool 1 LHAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LHAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI CGAP Pr22 pool 1 LHAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI CGAP Co10 pool 1 LHAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351). (6% of the driver population), plus a pool of 3,840 arrayed clones from NCI CGAP Sub1 (IMAGE Clonoids 2708616-2710535) and NCI CGAP Sub2 (IMAGE Clonoids 2710536-2712455) (4% of the driver population), plus a pool of 11,136 clones from NCI CGAP Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI CGAP Sub4 (IMAGE Clonoids 2723592-2729326) (40% of the driver population), plus a pool of 4032 clones from NCI CGAP Sub6 (IMAGE Clonoids 2728969-2733190) (40% of the driver population). Subtraction was performed as previously described (Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806. TAG_SEQ=None found"

ORIGIN

Query Match 70.0%; Score 18.2; DB 10; Length 486;
 Best Local Similarity 87.0%; Pred. No. 4.7e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACACGGCGAGCCCGAATCTAC 23
 Db 460 AACACGGAGAGTCAGAACTAC 438

RESULT 28
 CE189676
 LOCUS
 DEFINITION tigr-gss-dog-17000371445035 Dog Library Canis familiaris genomic,
 Genomic survey sequence.

ACCESSION CE189676
 VERSION CE189676.1 GI:35345329
 KEYWORDS GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (Bases 1 to 538)
 Kirchner, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.

TITLE The dog genome: survey sequencing and comparative analysis
 JOURNAL Science 301 (5641), 1898-1903 (2003)
 MEDLINE 14512627
 PUBMED 22875432

COMMENT Contact: Kirkness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

FEATURES
 source

1. 538
 Location/Qualifiers
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 70.0%; Score 18.2; DB 29; Length 538;
 Best Local Similarity 87.0%; Pred. No. 4.9e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACACGGCGAGCCCGAATCTAC 23
 Db 466 AACACGGCAAGTCTTAATCTAC 488

RESULT 29

CA182954/c
 LOCUS
 DEFINITION SCEZST3151A02.9 ST3 Saccharum officinarum cDNA clone SCEZST3151A02
 5', mRNA sequence.

ACCESSION CA182954
 VERSION CA182954.1 GI:35119239
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Saccharum.

1 (bases 1 to 703)
 Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
 The libraries that made SUCEST
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
 CONTACT: Arruda P
 Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089
 Email: parruda@unicamp.br
 Clone distribution: clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bccccenter.fcav.unesp.br
 Plate: 151 row: A column: 02
 Seq primer: T7 Promoter Primer.

FEATURES

source

1. 703
 Location/Qualifiers
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCEZST3151A02"
 /lab_host="DH10B"
 /clone_lib="ST3"
 /note="Organ: Fourth apical stalk internodes of adult plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from Fourth apical stalk internodes of adult plants]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at
 http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 70.0%; Score 18.2; DB 14; Length 703;
 Best Local Similarity 87.0%; Pred. No. 5.2e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AACGGCGAGCCCGAATCTACGA 25
 Db 272 AACGACCGAGCCCGAATCTACGA 250

RESULT 30

BE469759
 LOCUS

BE469759 735 bp mRNA linear EST 28-JUL-2000

DEFINITION IpHk02105 Head kidney cDNA library Ictalurus punctatus cDNA 5',
mRNA sequence.
ACCESSION BE469759
VERSION BE469759.1 GI:9560250
KEYWORDS EST
SOURCE Ictalurus punctatus (channel catfish)
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.
REFERENCE 1 (bases 1 to 735)
AUTHORS Cao, D., Kocabas, A., Ju, Z., Karsi, A., Li, P., Patterson, A. and
Liu, Z.J.
TITLE Transcriptome analysis of channel catfish (Ictalurus punctatus):
Gene cataloging and profiling from the head kidney
JOURNAL Unpublished (2000)
COMMENT Contact: Liu ZJ
The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell
and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: M13 Reverse
FEATURES
source
location/Qualifiers
1..735
/organism="Ictalurus punctatus"
/mol_type="mRNA"
/db_xref="taxon:7998"
/clone_lib="Head kidney cDNA library"
/note="Organ: Head kidney; Vector: pSport1; Site_1: NotI;
Site_2: SalI"
ORIGIN
Query Match 70.0%; Score 18.2; DB 10; Length 735;
Best Local Similarity 87.0%; Pred. No. 5.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 4 ACGGGAGGCCCGATCTACGAA 26
|||||
Db 591 ACGGCGTGGTCCGTATCTACGAA 613
Search completed: June 20, 2004, 14:14:45
Job time : 1240.98 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 28.7513 Seconds
(without alignments)

Sequence: 1 acaacgagggccgaatctacgaa 26
501.846 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacgagggccgaatctacgaa 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

Issued Patents NA:**

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq*

5: /cgn2_6/ptodata/2/ina/PCRU_COMB.seq*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| C 1 | 26 | 100.0 | 401 | 3 | US-08-990-823-75 |
| C 2 | 26 | 100.0 | 401 | 4 | US-09-477-135A-75 |
| C 3 | 26 | 100.0 | 4403765 | 3 | US-09-103-840A-2 |
| C 4 | 26 | 100.0 | 4411529 | 3 | US-09-103-840A-1 |
| C 5 | 18.6 | 71.5 | 505 | 4 | US-09-221-017B-376 |
| C 6 | 17.6 | 67.7 | 536165 | 4 | US-09-214-808-1 |
| C 7 | 17 | 65.4 | 725 | 3 | US-08-998-416-808 |
| C 8 | 17 | 65.4 | 879 | 4 | US-09-408-020-37 |
| C 9 | 17 | 65.4 | 1200 | 4 | US-09-252-991A-3224 |
| C 10 | 17 | 65.4 | 1203 | 4 | US-09-252-991A-3421 |
| C 11 | 17 | 65.4 | 1272 | 4 | US-09-252-991A-15250 |
| C 12 | 17 | 65.4 | 1614 | 4 | US-09-252-991A-15438 |
| C 13 | 17 | 65.4 | 1644 | 4 | US-09-252-991A-792 |
| C 14 | 17 | 65.4 | 1845 | 4 | US-09-252-991A-3262 |
| C 15 | 17 | 65.4 | 2907 | 4 | US-08-252-991A-831 |
| C 16 | 17 | 65.4 | 42432 | 4 | US-09-408-020-2 |
| C 17 | 16.6 | 63.8 | 762 | 4 | US-09-489-039A-2899 |
| C 18 | 16.6 | 63.8 | 1305 | 4 | US-09-489-039A-5991 |
| C 19 | 16.6 | 63.8 | 1350 | 4 | US-09-252-991A-1564 |
| C 20 | 16.6 | 63.8 | 1494 | 4 | US-09-252-991A-1456 |
| C 21 | 16.6 | 63.8 | 1536 | 4 | US-09-252-991A-1567 |
| C 22 | 16.4 | 63.1 | 438 | 4 | US-08-252-991A-15918 |
| C 23 | 16.4 | 63.1 | 546 | 4 | US-09-429-906B-1 |
| C 24 | 16.4 | 63.1 | 762 | 4 | US-09-252-991A-15888 |
| C 25 | 16.4 | 63.1 | 885 | 4 | US-09-252-991A-15774 |
| C 26 | 16.4 | 63.1 | 1068 | 4 | US-09-252-991A-15804 |
| C 27 | 16.2 | 62.3 | 3063 | 4 | US-09-252-991A-12299 |

| | | | | | |
|-------|------|------|---------|---|----------------------|
| 28 | 16.2 | 62.3 | 3105 | 4 | US-09-252-991A-12350 |
| C 29 | 16.2 | 62.3 | 4403765 | 3 | US-09-103-840A-2 |
| C 30 | 16.2 | 62.3 | 4411529 | 3 | US-09-103-840A-1 |
| C 31 | 16 | 61.5 | 420 | 1 | US-08-470-179-148 |
| C 32 | 16 | 61.5 | 489 | 4 | US-09-894-844-63 |
| C 33 | 16 | 61.5 | 696 | 4 | US-09-252-991A-12599 |
| C 34 | 16 | 61.5 | 971 | 4 | US-08-630-915A-197 |
| C 35 | 16 | 61.5 | 1457 | 3 | US-09-444-053-3 |
| C 36 | 16 | 61.5 | 4079 | 4 | US-09-016-434-1248 |
| C 37 | 16 | 61.5 | 4519 | 4 | US-09-023-655-1202 |
| C 38 | 16 | 61.5 | 15378 | 3 | US-08-785-420-1 |
| C 39 | 15.6 | 60.0 | 423 | 1 | US-08-470-179-107 |
| C 40 | 15.6 | 60.0 | 423 | 1 | US-08-470-179-133 |
| C 41 | 15.6 | 60.0 | 423 | 1 | US-08-470-179-151 |
| C 42 | 15.6 | 60.0 | 426 | 1 | US-08-470-179-179 |
| C 43 | 15.6 | 60.0 | 456 | 4 | US-09-325-932A-14 |
| C 44 | 15.6 | 60.0 | 861 | 4 | US-09-252-991A-15703 |
| C 45 | 15.6 | 60.0 | 891 | 4 | US-09-252-991A-15676 |
| C 46 | 15.6 | 60.0 | 11050 | 4 | US-10-204-708-85 |
| C 47 | 15.4 | 59.2 | 294 | 1 | US-08-221-817-9 |
| C 48 | 15.4 | 59.2 | 294 | 1 | US-08-454-439-9 |
| C 49 | 15.4 | 59.2 | 294 | 4 | US-09-107-532A-217 |
| C 50 | 15.4 | 59.2 | 294 | 5 | PCT-US94-10487-9 |
| C 51 | 15.4 | 59.2 | 346 | 3 | US-08-998-416-68 |
| C 52 | 15.4 | 59.2 | 346 | 3 | US-08-998-416-71 |
| C 53 | 15.4 | 59.2 | 360 | 4 | US-09-252-991A-15436 |
| C 54 | 15.4 | 59.2 | 474 | 4 | US-09-252-991A-6449 |
| C 55 | 15.4 | 59.2 | 531 | 4 | US-09-252-991A-8239 |
| C 56 | 15.4 | 59.2 | 642 | 4 | US-09-252-991A-12174 |
| C 57 | 15.4 | 59.2 | 879 | 4 | US-09-408-020-71 |
| C 58 | 15.4 | 59.2 | 1059 | 4 | US-09-489-039A-3376 |
| C 59 | 15.4 | 59.2 | 1059 | 4 | US-09-791-165-5 |
| C 60 | 15.4 | 59.2 | 1059 | 4 | US-09-791-165-6 |
| C 61 | 15.4 | 59.2 | 1104 | 4 | US-09-489-039A-198 |
| C 62 | 15.4 | 59.2 | 1239 | 4 | US-09-252-991A-8289 |
| C 63 | 15.4 | 59.2 | 1314 | 4 | US-09-495-050A-174 |
| C 64 | 15.4 | 59.2 | 1335 | 4 | US-09-252-991A-12238 |
| C 65 | 15.4 | 59.2 | 1360 | 4 | US-09-602-472A-14 |
| C 66 | 15.4 | 59.2 | 1671 | 4 | US-09-252-991A-6099 |
| C 67 | 15.4 | 59.2 | 1674 | 4 | US-09-252-991A-6179 |
| C 68 | 15.4 | 59.2 | 1688 | 4 | US-09-791-165-3 |
| C 69 | 15.4 | 59.2 | 1700 | 4 | US-09-514-521-2 |
| C 70 | 15.4 | 59.2 | 1700 | 4 | US-09-791-165-1 |
| C 71 | 15.4 | 59.2 | 1743 | 4 | US-09-252-991A-6372 |
| C 72 | 15.4 | 59.2 | 1887 | 4 | US-09-149-727-9 |
| C 73 | 15.4 | 59.2 | 1931 | 4 | US-09-429-906B-5 |
| C 74 | 15.4 | 59.2 | 2364 | 4 | US-09-252-991A-12368 |
| C 75 | 15.4 | 59.2 | 2639 | 1 | US-07-952-817-8 |
| C 76 | 15.4 | 59.2 | 2639 | 6 | 5210025-1 |
| C 77 | 15.4 | 59.2 | 3456 | 4 | US-09-602-472A-12 |
| C 78 | 15.4 | 59.2 | 3556 | 1 | US-07-971-624E-1 |
| C 79 | 15.4 | 59.2 | 3572 | 1 | US-07-971-624E-2 |
| C 80 | 15.4 | 59.2 | 3660 | 4 | US-09-107-532A-2366 |
| C 81 | 15.4 | 59.2 | 3843 | 4 | US-09-252-991A-1430 |
| C 82 | 15.4 | 59.2 | 3931 | 4 | US-08-125-468-1 |
| C 83 | 15.4 | 59.2 | 3931 | 2 | US-08-474-933-1 |
| C 84 | 15.4 | 59.2 | 3931 | 2 | US-09-408-020-1 |
| C 85 | 15.4 | 59.2 | 3939 | 4 | US-09-408-020-1 |
| C 86 | 15.4 | 59.2 | 3939 | 4 | US-08-752-760A-1 |
| C 87 | 15.2 | 58.5 | 346 | 3 | US-09-602-472A-12 |
| C 88 | 15.2 | 58.5 | 346 | 4 | US-09-670-756-317 |
| C 89 | 15.2 | 58.5 | 676 | 4 | US-09-470-191-49 |
| C 90 | 15.2 | 58.5 | 942 | 4 | US-09-252-991A-10909 |
| C 91 | 15.2 | 58.5 | 945 | 4 | US-09-252-991A-10966 |
| C 92 | 15.2 | 58.5 | 1278 | 4 | US-09-252-991A-11277 |
| C 93 | 15.2 | 58.5 | 1806 | 4 | US-09-252-991A-11213 |
| C 94 | 15.2 | 58.5 | 2312 | 1 | US-07-736-178C-1 |
| C 95 | 15.2 | 58.5 | 2325 | 2 | US-08-714-677-3 |
| C 96 | 15.2 | 58.5 | 2325 | 2 | US-08-393-540-3 |
| C 97 | 15.2 | 58.5 | 2325 | 2 | US-08-714-537-3 |
| C 98 | 15.2 | 58.5 | 3073 | 2 | US-08-714-677-11 |
| C 99 | 15.2 | 58.5 | 3073 | 2 | US-08-393-540-11 |
| C 100 | 15.2 | 58.5 | 3073 | 2 | US-08-714-537-11 |

Sequence 12350, A
Sequence 2, Appli
Sequence 1, Appli
Sequence 148, Appl
Sequence 63, Appl
Sequence 12599, A
Sequence 197, Appl
Sequence 3, Appli
Sequence 1248, Ap
Sequence 1202, Ap
Sequence 1, Appli
Sequence 107, Appl
Sequence 133, Appl
Sequence 151, Appl
Sequence 179, Appl
Sequence 14, Appl
Sequence 15703, A
Sequence 15676, A
Sequence 85, Appl
Sequence 9, Appli
Sequence 9, Appli
Sequence 217, Appl
Sequence 9, Appli
Sequence 68, Appl
Sequence 71, Appl
Sequence 15436, A
Sequence 6449, Ap
Sequence 8239, Ap
Sequence 12174, A
Sequence 71, Appl
Sequence 3376, Ap
Sequence 5, Appli
Sequence 6, Appli
Sequence 198, Appl
Sequence 8289, Ap
Sequence 174, Appl
Sequence 12338, A
Sequence 14, Appl
Sequence 6099, Ap
Sequence 6179, Ap
Sequence 3, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 6372, Ap
Sequence 5, Appli
Sequence 5, Appli
Sequence 12368, A
Sequence 8, Appli
Patent No. 5210025
Sequence 12, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 2366, Ap
Sequence 1430, Ap
Sequence 1588, Ap
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 317, Appl
Sequence 317, Appl
Sequence 49, Appl
Sequence 10309, A
Sequence 10566, A
Sequence 11277, A
Sequence 11213, A
Sequence 1, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 11, Appl
Sequence 11, Appl
Sequence 11, Appl

| | | | | | | | | | | | | | |
|-----|------|------|-------|---|----------------------|--------------------|-------|------|------|-------|---|----------------------|--------------------|
| 101 | 15.2 | 58.5 | 7685 | 4 | US-09-502-710-22 | Sequence 22, Appl | c 174 | 15 | 57.7 | 11233 | 3 | US-08-980-832-27 | Sequence 27, Appl |
| 102 | 15.2 | 58.5 | 7685 | 4 | US-09-502-710-25 | Sequence 25, Appl | c 175 | 15 | 57.7 | 11233 | 4 | US-09-920-923B-27 | Sequence 27, Appl |
| 103 | 15.2 | 58.5 | 7685 | 4 | US-09-502-711-22 | Sequence 22, Appl | 176 | 15 | 57.7 | 30001 | 1 | US-08-125-468-1 | Sequence 1, Appl |
| 104 | 15.2 | 58.5 | 7685 | 4 | US-09-502-711-25 | Sequence 25, Appl | 177 | 15 | 57.7 | 30001 | 2 | US-08-474-933-1 | Sequence 1, Appl |
| 105 | 15.2 | 58.5 | 7685 | 4 | US-09-555-616A-1 | Sequence 1, Appl | 178 | 15 | 57.7 | 36412 | 4 | US-08-311-731A-132 | Sequence 132, App |
| 106 | 15.2 | 58.5 | 7686 | 4 | US-09-502-710-23 | Sequence 23, Appl | c 179 | 15 | 57.7 | 45716 | 4 | US-08-965-048-5 | Sequence 5, Appl |
| 107 | 15.2 | 58.5 | 7686 | 4 | US-09-502-710-26 | Sequence 26, Appl | c 180 | 15 | 57.7 | 45989 | 4 | US-08-965-048-6 | Sequence 6, Appl |
| 108 | 15.2 | 58.5 | 7686 | 4 | US-09-502-711-23 | Sequence 23, Appl | 181 | 14.8 | 56.9 | 259 | 4 | US-08-621-976-12928 | Sequence 12928, A |
| 109 | 15.2 | 58.5 | 7686 | 4 | US-09-502-711-26 | Sequence 26, Appl | c 182 | 14.8 | 56.9 | 372 | 4 | US-09-134-001C-425 | Sequence 425, App |
| 110 | 15.2 | 58.5 | 7686 | 4 | US-09-502-710-24 | Sequence 24, Appl | 183 | 14.8 | 56.9 | 413 | 4 | US-08-621-976-14233 | Sequence 14233, A |
| 111 | 15.2 | 58.5 | 7687 | 4 | US-09-502-711-24 | Sequence 24, Appl | 184 | 14.8 | 56.9 | 423 | 4 | US-08-252-991A-1756 | Sequence 1756, App |
| 112 | 15.2 | 58.5 | 7687 | 4 | US-09-502-711-27 | Sequence 27, Appl | 185 | 14.8 | 56.9 | 432 | 4 | US-08-833-381-600 | Sequence 600, App |
| 113 | 15.2 | 58.5 | 7688 | 4 | US-09-502-710-27 | Sequence 27, Appl | 186 | 14.8 | 56.9 | 439 | 4 | US-08-621-976-14007 | Sequence 14007, A |
| 114 | 15.2 | 58.5 | 7688 | 4 | US-09-221-017B-636 | Sequence 636, App | 187 | 14.8 | 56.9 | 447 | 4 | US-08-621-976-14007 | Sequence 14007, A |
| 115 | 15.2 | 58.5 | 9837 | 4 | US-09-410-551B-1 | Sequence 1, Appl | 188 | 14.8 | 56.9 | 458 | 4 | US-08-621-976-12510 | Sequence 12510, A |
| 116 | 15.2 | 58.5 | 77536 | 4 | US-09-410-551B-1 | Sequence 1, Appl | 189 | 14.8 | 56.9 | 465 | 4 | US-08-621-976-10666 | Sequence 10666, A |
| 117 | 15.2 | 58.5 | 77536 | 4 | US-09-621-976-18428 | Sequence 18428, A | 190 | 14.8 | 56.9 | 477 | 4 | US-08-621-976-10666 | Sequence 10666, A |
| 118 | 15 | 57.7 | 186 | 4 | 5171672-2 | Patent No. 5171672 | 191 | 14.8 | 56.9 | 480 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 119 | 15 | 57.7 | 352 | 6 | US-08-928-213B-180 | Sequence 180, App | 192 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 120 | 15 | 57.7 | 444 | 3 | US-08-928-213B-180 | Sequence 180, App | 193 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 121 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 194 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 122 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 195 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 123 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 196 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 124 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 197 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 125 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 198 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 126 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 199 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 127 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 200 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 128 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 201 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 129 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 202 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 130 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 203 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 131 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 204 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 132 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 205 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 133 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 206 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 134 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 207 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 135 | 15 | 57.7 | 485 | 2 | US-08-928-213B-180 | Sequence 180, App | 208 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 136 | 15 | 57.7 | 522 | 4 | US-08-928-213B-180 | Sequence 180, App | 209 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 137 | 15 | 57.7 | 660 | 4 | US-08-928-213B-180 | Sequence 180, App | 210 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 138 | 15 | 57.7 | 966 | 4 | US-08-928-213B-180 | Sequence 180, App | 211 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 139 | 15 | 57.7 | 1062 | 4 | US-08-928-213B-180 | Sequence 180, App | 212 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 140 | 15 | 57.7 | 1113 | 4 | US-08-928-213B-180 | Sequence 180, App | 213 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 141 | 15 | 57.7 | 1155 | 4 | US-08-928-213B-180 | Sequence 180, App | 214 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 142 | 15 | 57.7 | 1158 | 4 | US-08-928-213B-180 | Sequence 180, App | 215 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 143 | 15 | 57.7 | 1281 | 4 | US-08-928-213B-180 | Sequence 180, App | 216 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 144 | 15 | 57.7 | 1341 | 4 | US-08-928-213B-180 | Sequence 180, App | 217 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 145 | 15 | 57.7 | 1392 | 4 | US-08-928-213B-180 | Sequence 180, App | 218 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 146 | 15 | 57.7 | 1740 | 4 | US-08-928-213B-180 | Sequence 180, App | 219 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 147 | 15 | 57.7 | 1746 | 4 | US-08-928-213B-180 | Sequence 180, App | 220 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 148 | 15 | 57.7 | 1785 | 4 | US-08-928-213B-180 | Sequence 180, App | 221 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 149 | 15 | 57.7 | 2025 | 4 | US-08-928-213B-180 | Sequence 180, App | 222 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 150 | 15 | 57.7 | 2028 | 4 | US-08-928-213B-180 | Sequence 180, App | 223 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 151 | 15 | 57.7 | 2082 | 4 | US-08-928-213B-180 | Sequence 180, App | 224 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 152 | 15 | 57.7 | 2094 | 4 | US-08-928-213B-180 | Sequence 180, App | 225 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 153 | 15 | 57.7 | 2100 | 4 | US-08-928-213B-180 | Sequence 180, App | 226 | 14.8 | 56.9 | 481 | 4 | US-08-621-976-10533 | Sequence 10533, A |
| 154 | 15 | 57.7 | 2220 | 6 | 5171850-1 | Patent No. 5171850 | 227 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-8 | Sequence 8, Appl |
| 155 | 15 | 57.7 | 2409 | 4 | US-09-252-991A-8479 | Sequence 8479, App | 228 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-9 | Sequence 9, Appl |
| 156 | 15 | 57.7 | 2600 | 4 | US-09-252-991A-8479 | Sequence 8479, App | 229 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-163 | Sequence 163, App |
| 157 | 15 | 57.7 | 2905 | 4 | US-09-468-578-3 | Sequence 3, Appl | 230 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-164 | Sequence 164, App |
| 158 | 15 | 57.7 | 2905 | 4 | US-09-468-578-3 | Sequence 3, Appl | 231 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-182 | Sequence 182, App |
| 159 | 15 | 57.7 | 2905 | 4 | US-09-468-578-3 | Sequence 3, Appl | 232 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-183 | Sequence 183, App |
| 160 | 15 | 57.7 | 2970 | 4 | US-09-252-991A-15898 | Sequence 15898, A | 233 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-174 | Sequence 174, App |
| 161 | 15 | 57.7 | 3176 | 1 | US-08-425-239A-3 | Sequence 3, Appl | 234 | 14.6 | 56.2 | 63 | 1 | US-08-086-630C-166 | Sequence 166, App |
| 162 | 15 | 57.7 | 3176 | 1 | US-08-425-239A-3 | Sequence 3, Appl | 235 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-6312 | Sequence 6312, App |
| 163 | 15 | 57.7 | 3176 | 1 | US-08-425-239A-3 | Sequence 3, Appl | 236 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-14924 | Sequence 14924, A |
| 164 | 15 | 57.7 | 3177 | 2 | US-09-070-060-10 | Sequence 10, Appl | 237 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-14924 | Sequence 14924, A |
| 165 | 15 | 57.7 | 3177 | 3 | US-09-357-746-10 | Sequence 10, Appl | 238 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 166 | 15 | 57.7 | 3267 | 3 | US-08-633-768A-3 | Sequence 3, Appl | 239 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 167 | 15 | 57.7 | 3267 | 3 | US-08-633-768A-3 | Sequence 3, Appl | 240 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 168 | 15 | 57.7 | 3486 | 4 | US-09-543-681A-2314 | Sequence 2314, App | 241 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 169 | 15 | 57.7 | 3836 | 2 | US-08-218-260-1 | Sequence 1, Appl | 242 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 170 | 15 | 57.7 | 3956 | 2 | US-08-966-388-9 | Sequence 9, Appl | 243 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 171 | 15 | 57.7 | 3956 | 3 | US-09-188-403-9 | Sequence 9, Appl | 244 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 172 | 15 | 57.7 | 3956 | 3 | US-09-188-403-9 | Sequence 9, Appl | 245 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |
| 173 | 15 | 57.7 | 3956 | 3 | US-09-281-259-9 | Sequence 9, Appl | 246 | 14.6 | 56.2 | 63 | 1 | US-09-252-991A-1585 | Sequence 1585, App |

C 247 14.6 56.2 1458 4 US-09-252-991A-14586
 C 248 14.6 56.2 1551 4 US-09-489-039A-4788
 C 249 14.6 56.2 1674 4 US-09-252-991A-11512
 C 250 14.6 56.2 1845 4 US-09-252-991A-8187
 C 251 14.6 56.2 2074 4 US-09-690-942-7
 C 252 14.6 56.2 2167 3 US-08-743-637B-16
 C 253 14.6 56.2 2167 3 US-08-526-840B-16
 C 254 14.6 56.2 2310 3 US-09-036-987A-25
 C 255 14.6 56.2 2310 3 US-09-370-700-25
 C 256 14.6 56.2 2310 3 US-09-603-207-25
 C 257 14.6 56.2 2367 4 US-09-489-039A-5660
 C 258 14.6 56.2 2457 4 US-09-252-991A-15181
 C 259 14.6 56.2 3098 4 US-09-252-991A-7487
 C 260 14.6 56.2 3938 4 US-10-164-595-45
 C 261 14.6 56.2 4131 4 US-09-252-991A-12944
 C 262 14.6 56.2 4326 4 US-09-252-991A-11572
 C 263 14.6 56.2 4378 4 US-10-164-595-47
 C 264 14.6 56.2 4473 4 US-09-252-991A-11871
 C 265 14.6 56.2 4648 4 US-09-620-312D-464
 C 266 14.6 56.2 6131 1 US-07-732-242C-8
 C 267 14.6 56.2 6051 2 US-08-576-626A-2
 C 268 14.6 56.2 9775 4 US-09-647-540A-1
 C 269 14.6 56.2 9775 4 US-09-647-540A-3
 C 270 14.6 56.2 9775 4 US-09-647-540A-5
 C 271 14.6 56.2 9775 4 US-09-647-540A-7
 C 272 14.6 56.2 9775 4 US-10-119-600-1
 C 273 14.6 56.2 9775 4 US-10-119-600-3
 C 274 14.6 56.2 9775 4 US-10-119-600-5
 C 275 14.6 56.2 9775 4 US-10-119-600-7
 C 276 14.6 56.2 9775 4 US-10-119-651-1
 C 277 14.6 56.2 9775 4 US-10-119-651-3
 C 278 14.6 56.2 9775 4 US-10-119-651-5
 C 279 14.6 56.2 9775 4 US-10-119-651-7
 C 280 14.6 56.2 34094 4 US-09-292-034-1
 C 281 14.6 56.2 36412 4 US-08-311-731A-132
 C 282 14.4 55.4 28 2 US-08-859-988-732
 C 283 14.4 55.4 28 4 US-09-225-328-732
 C 284 14.4 55.4 28 4 US-09-225-201B-732
 C 285 14.4 55.4 110 4 US-09-621-976-9864
 C 286 14.4 55.4 186 4 US-09-313-294A-5677
 C 287 14.4 55.4 270 4 US-09-489-039A-1546
 C 288 14.4 55.4 273 4 US-09-252-991A-4560
 C 289 14.4 55.4 348 3 US-09-060-756-142
 C 290 14.4 55.4 348 4 US-09-670-314-142
 C 291 14.4 55.4 363 4 US-09-489-039A-3412
 C 292 14.4 55.4 371 3 US-09-060-756-199
 C 293 14.4 55.4 371 4 US-09-670-314-199
 C 294 14.4 55.4 396 1 US-08-470-179-181
 C 295 14.4 55.4 420 1 US-08-470-179-162
 C 296 14.4 55.4 423 1 US-08-470-179-175
 C 297 14.4 55.4 423 1 US-08-470-179-186
 C 298 14.4 55.4 423 1 US-08-470-179-189
 C 299 14.4 55.4 443 3 US-09-060-756-172
 C 300 14.4 55.4 443 4 US-09-670-314-172

ALIGNMENTS

RESULT 1
 US-08-990-823-75/c
 ; Sequence 75, Application US/08990823D
 ; Patent No. 6228371
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: Immunostimulatory Peptides
 ; FILE REFERENCE: 49086
 ; CURRENT APPLICATION NUMBER: US/08/990,823D
 ; CURRENT FILING DATE: 1997-12-15
 ; EARLIER APPLICATION NUMBER: US 96/10375
 ; EARLIER FILING DATE: 1996-06-14
 ; EARLIER APPLICATION NUMBER: 60/000,254

; EARLIER FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 113
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 75
 ; LENGTH: 401
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; NAME/KEY: Modified base
 ; OTHER INFORMATION: n represents a o r g o r c o r t/u
 US-08-990-823-75
 Query Match 100.0%; Score 26; DB 3; Length 401;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCCCGCAATCTACGAA 26
 Db 148 ACAACGGCGAGCCCGCAATCTACGAA 123
 RESULT 2
 US-09-477-135A-75/c
 ; Sequence 75, Application US/09477135A
 ; Patent No. 6572865
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: Immunostimulatory Peptides
 ; FILE REFERENCE: 52888
 ; CURRENT APPLICATION NUMBER: US/09/477,135A
 ; CURRENT FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08990823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 75
 ; LENGTH: 401
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; NAME/KEY: variation
 ; LOCATION: (1)..(401)
 ; OTHER INFORMATION: n = A, C, G, or T
 US-09-477-135A-75
 Query Match 100.0%; Score 26; DB 4; Length 401;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCCCGCAATCTACGAA 26
 Db 148 ACAACGGCGAGCCCGCAATCTACGAA 123
 RESULT 3
 US-09-103-840A-2
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6234328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 26; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26
DB 4219170 ACAACGGCGAGCCCGGAATCTACGAA 4219195

RESULT 4
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 100.0%; Score 26; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26
DB 4226924 ACAACGGCGAGCCCGGAATCTACGAA 4226949

RESULT 5
US-09-221-017B-376
Sequence 376, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b

Query Match 100.0%; Score 26; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26
DB 4226924 ACAACGGCGAGCCCGGAATCTACGAA 4226949

RESULT 6
US-09-214-808-1/c
Sequence 1, Application US/09214808A
Patent No. 6475793
GENERAL INFORMATION:
APPLICANT: Rosenthal, Andre
APPLICANT: Freiberg, Christoph
APPLICANT: Perret, Xavier Philippe
APPLICANT: Broughton, William John
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
Patent No. 6475793
TITLE OF INVENTION: Plasmid
FILE REFERENCE: CARP0068
CURRENT APPLICATION NUMBER: US/09/214,808A
CURRENT FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: PCT/IB97/00950
PRIOR FILING DATE: 1997-07-10
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 1
LENGTH: 536165
TYPE: DNA
ORGANISM: Rhizobium
US-09-214-808-1

Query Match 71.5%; Score 18.6; DB 4; Length 505;
Best Local Similarity 84.0%; Pred. No. 7.3;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGGAATCTACGAA 26
DB 336 CAACGGCGATGCGCGGAATCAACGAA 360

RESULT 6
US-09-214-808-1/c
Sequence 1, Application US/09214808A
Patent No. 6475793
GENERAL INFORMATION:
APPLICANT: Rosenthal, Andre
APPLICANT: Freiberg, Christoph
APPLICANT: Perret, Xavier Philippe
APPLICANT: Broughton, William John
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
Patent No. 6475793
TITLE OF INVENTION: Plasmid
FILE REFERENCE: CARP0068
CURRENT APPLICATION NUMBER: US/09/214,808A
CURRENT FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: PCT/IB97/00950
PRIOR FILING DATE: 1997-07-10
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 1
LENGTH: 536165
TYPE: DNA
ORGANISM: Rhizobium
US-09-214-808-1

Query Match 71.5%; Score 18.6; DB 4; Length 505;
Best Local Similarity 84.0%; Pred. No. 7.3;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGGAATCTACGAA 26
DB 336 CAACGGCGATGCGCGGAATCAACGAA 360

Query Match 67.7%; Score 17.6; DB 4; Length 536165;
Best Local Similarity 83.3%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGCCCGAATCTACGAA 26
85320 AAGGGCGGCGCCCGAACAACGAA 85297

RESULT. T 7

```

RESOL:
US-08L-998-416-808
/ Sequence 808, Application US/08998416
/ Patent No. 6239264
/ GENERAL INFORMATION:
/ APPLICANT: Philippssen, Peter
/ APPLICANT: Pohlmann, Rainer
/ APPLICANT: Steiner, Sabine
/ APPLICANT: Mohr, Christine
/ APPLICANT: Wendland, Jurgin
/ APPLICANT: Knechtle, Philipp
/ APPLICANT: Reibschung, Corinne
/ TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYIPII
/ TITLE OF INVENTION: AND USES THEREOF
/ NUMBER OF SEQUENCES: 1152

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 808:

```

/ INFORMATION FOR SEQ ID NO: 808:
/
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 725 base pairs
/   TYPE: nucleic acid
/   STRANDEDNESS: single
/   TOPOLOGY: linear
/
/ MOLSCUE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: PAG1515UP
/ US-08-999-416-908

```

Query Match 65.4%; Score 17; DB 3; Length 725;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 5; Indels

Qy 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 472 ACAACGGCGAGGCATGCATGTACGA 496

RESULT 8
US-09-408-020-37/c

```

1  Patent No. 6632937
2  /
3  /
4  /
5  /
6  /
7  /
8  /
9  /
10 /
11 /
12 /
13 /
14 /
15 /
16 /
17 /
18 /
19 /
20 /
21 /
22 /
23 /
24 /
25 /
26 /
27 /
28 /
29 /
30 /
31 /
32 /
33 /
34 /
35 /
36 /
37 /
38 /
39 /
40 /
41 /
42 /
43 /
44 /
45 /
46 /
47 /
48 /
49 /
50 /
51 /
52 /
53 /
54 /
55 /
56 /
57 /
58 /
59 /
60 /
61 /
62 /
63 /
64 /
65 /
66 /
67 /
68 /
69 /
70 /
71 /
72 /
73 /
74 /
75 /
76 /
77 /
78 /
79 /
80 /
81 /
82 /
83 /
84 /
85 /
86 /
87 /
88 /
89 /
90 /
91 /
92 /
93 /
94 /
95 /
96 /
97 /
98 /
99 /
100 /
101 /
102 /
103 /
104 /
105 /
106 /
107 /
108 /
109 /
110 /
111 /
112 /
113 /
114 /
115 /
116 /
117 /
118 /
119 /
120 /
121 /
122 /
123 /
124 /
125 /
126 /
127 /
128 /
129 /
130 /
131 /
132 /
133 /
134 /
135 /
136 /
137 /
138 /
139 /
140 /
141 /
142 /
143 /
144 /
145 /
146 /
147 /
148 /
149 /
150 /
151 /
152 /
153 /
154 /
155 /
156 /
157 /
158 /
159 /
160 /
161 /
162 /
163 /
164 /
165 /
166 /
167 /
168 /
169 /
170 /
171 /
172 /
173 /
174 /
175 /
176 /
177 /
178 /
179 /
180 /
181 /
182 /
183 /
184 /
185 /
186 /
187 /
188 /
189 /
190 /
191 /
192 /
193 /
194 /
195 /
196 /
197 /
198 /
199 /
200 /
201 /
202 /
203 /
204 /
205 /
206 /
207 /
208 /
209 /
210 /
211 /
212 /
213 /
214 /
215 /
216 /
217 /
218 /
219 /
220 /
221 /
222 /
223 /
224 /
225 /
226 /
227 /
228 /
229 /
230 /
231 /
232 /
233 /
234 /
235 /
236 /
237 /
238 /
239 /
240 /
241 /
242 /
243 /
244 /
245 /
246 /
247 /
248 /
249 /
250 /
251 /
252 /
253 /
254 /
255 /
256 /
257 /
258 /
259 /
260 /
261 /
262 /
263 /
264 /
265 /
266 /
267 /
268 /
269 /
270 /
271 /
272 /
273 /
274 /
275 /
276 /
277 /
278 /
279 /
280 /
281 /
282 /
283 /
284 /
285 /
286 /
287 /
288 /
289 /
290 /
291 /
292 /
293 /
294 /
295 /
296 /
297 /
298 /
299 /
300 /
301 /
302 /
303 /
304 /
305 /
306 /
307 /
308 /
309 /
310 /
311 /
312 /
313 /
314 /
315 /
316 /
317 /
318 /
319 /
320 /
321 /
322 /
323 /
324 /
325 /
326 /
327 /
328 /
329 /
330 /
331 /
332 /
333 /
334 /
335 /
336 /
337 /
338 /
339 /
340 /
341 /
342 /
343 /
344 /
345 /
346 /
347 /
348 /
349 /
350 /
351 /
352 /
353 /
354 /
355 /
356 /
357 /
358 /
359 /
360 /
361 /
362 /
363 /
364 /
365 /
366 /
367 /
368 /
369 /
370 /
371 /
372 /
373 /
374 /
375 /
376 /
377 /
378 /
379 /
380 /
381 /
382 /
383 /
384 /
385 /
386 /
387 /
388 /
389 /
390 /
391 /
392 /
393 /
394 /
395 /
396 /
397 /
398 /
399 /
400 /
401 /
402 /
403 /
404 /
405 /
406 /
407 /
408 /
409 /
410 /
411 /
412 /
413 /
414 /
415 /
416 /
417 /
418 /
419 /
420 /
421 /
422 /
423 /
424 /
425 /
426 /
427 /
428 /
429 /
430 /
431 /
432 /
433 /
434 /
435 /
436 /
437 /
438 /
439 /
440 /
441 /
442 /
443 /
444 /
445 /
446 /
447 /
448 /
449 /
450 /
451 /
452 /
453 /
454 /
455 /
456 /
457 /
458 /
459 /
460 /
461 /
462 /
463 /
464 /
465 /
466 /
467 /
468 /
469 /
470 /
471 /
472 /
473 /
474 /
475 /
476 /
477 /
478 /
479 /
480 /
481 /
482 /
483 /
484 /
485 /
486 /
487 /
488 /
489 /
490 /
491 /
492 /
493 /
494 /
495 /
496 /
497 /
498 /
499 /
500 /
501 /
502 /
503 /
504 /
505 /
506 /
507 /
508 /
509 /
510 /
511 /
512 /
513 /
514 /
515 /
516 /
517 /
518 /
519 /
520 /
521 /
522 /
523 /
524 /
525 /
526 /
527 /
528 /
529 /
530 /
531 /
532 /
533 /
534 /
535 /
536 /
537 /
538 /
539 /
540 /
541 /
542 /
543 /
544 /
545 /
546 /
547 /
548 /
549 /
550 /
551 /
552 /
553 /
554 /
555 /
556 /
557 /
558 /
559 /
560 /
561 /
562 /
563 /
564 /
565 /
566 /
567 /
568 /
569 /
570 /
571 /
572 /
573 /
574 /
575 /
576 /
577 /
578 /
579 /
580 /
581 /
582 /
583 /
584 /
585 /
586 /
587 /
588 /
589 /
590 /
591 /
592 /
593 /
594 /
595 /
596 /
597 /
598 /
599 /
600 /
601 /
602 /
603 /
604 /
605 /
606 /
607 /
608 /
609 /
610 /
611 /
612 /
613 /
614 /
615 /
616 /
617 /
618 /
619 /
620 /
621 /
622 /
623 /
624 /
625 /
626 /
627 /
628 /
629 /
630 /
631 /
632 /
633 /
634 /
635 /
636 /
637 /
638 /
639 /
640 /
641 /
642 /
643 /
644 /
645 /
646 /
647 /
648 /
649 /
650 /
651 /
652 /
653 /
654 /
655 /
656 /
657 /
658 /
659 /
660 /
661 /
662 /
663 /
664 /
665 /
666 /
667 /
668 /
669 /
670 /
671 /
672 /
673 /
674 /
675 /
676 /
677 /
678 /
679 /
680 /
681 /
682 /
683 /
684 /
685 /
686 /
687 /
688 /
689 /
690 /
691 /
692 /
693 /
694 /
695 /
696 /
697 /
698 /
699 /
700 /
701 /
702 /
703 /
704 /
705 /
706 /
707 /
708 /
709 /
710 /
711 /
712 /
713 /
714 /
715 /
716 /
717 /
718 /
719 /
720 /
721 /
722 /
723 /
724 /
725 /
726 /
727 /
728 /
729 /
730 /
731 /
732 /
733 /
734 /
735 /
736 /
737 /
738 /
739 /
740 /
741 /
742 /
743 /
744 /
745 /
746 /
747 /
748 /
749 /
750 /
751 /
752 /
753 /
754 /
755 /
756 /
757 /
758 /
759 /
760 /
761 /
762 /
763 /
764 /
765 /
766 /
767 /
768 /
769 /
770 /
771 /
772 /
773 /
774 /
775 /
776 /
777 /
778 /
779 /
780 /
781 /
782 /
783 /
784 /
785 /
786 /
787 /
788 /
789 /
790 /
791 /
792 /
793 /
794 /
795 /
796 /
797 /
798 /
799 /
800 /
801 /
802 /
803 /
804 /
805 /
806 /
807 /
808 /
809 /
810 /
811 /
812 /
813 /
814 /
815 /
816 /
817 /
818 /
819 /
820 /
821 /
822 /
823 /
824 /
825 /
826 /
827 /
828 /
829 /
830 /
831 /
832 /
833 /
834 /
835 /
836 /
83
```

```

; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 879

```

```

; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(879)
US-09-408-020-37

```

| | | | | |
|-----------------------|--------------|---------------|------------|-------------|
| Query Match | 65.4% | Score 17; | DB 4; | Length 879; |
| Best Local Similarity | 80.0% | Pred. No. 44; | | |
| Matches: 20; | Conservative | 0; | Mismatches | 5; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

QY 1 ACAACGGCGAGGCCCGAATCTACGA 25
434 ACCTCGGAGAGGCCCGAGTCTACAA 410
Db

RESULT 9

```

US-09-252-991A-3224
; Sequence 3224, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: AERUGINOSA FOR DI
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,7
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,1
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3224
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3224

```

Query Match 65.4%; Score 17; DB 4; Length 1200;
Best Local Similarity 80.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 5; Indels

Qy 2 CAACGGCGAGGCCGGAATCTACGAA 26
Dp 925 CAGCGGCGAGGCCGCGTTCTACGAA 949

RESULT. T 10

RES-09-252-91A-3421/C
 Sequence 3421, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND
 AERUGINOSA FOR DIAGNOSIS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,


```

US-09-252-991A-831
; Sequence 831, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 831
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-831

Query Match      65.4%; Score 17; DB 4; Length 2907;
Best Local Similarity 80.0%; Pred. No. 48;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 ACAACGGCGAGGCCCGCAATCTACGA 25
Db      812 ACAACGGCGAGGCCCGCAATCTACGA 836

RESULT 16
US-09-408-020-2/c
; Sequence 2, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOEP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11478)...(13046)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: (29655)....(30491)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34559)....(36067)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37002)....(37403)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37404)....(38282)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (39454)....(40572)
US-09-408-020-2

Query Match      65.4%; Score 17; DB 4; Length 42432;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 ACAACGGCGAGGCCCGCAATCTACGA 25
Db      37837 ACCTCGGAGAGGCCCGGAGTCTACAA 37813

RESULT 17
US-09-489-039A-2899
; Sequence 2899, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2899
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2899

Query Match      63.8%; Score 16.6; DB 4; Length 762;
Best Local Similarity 82.6%; Pred. No. 66;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CAACGGCGAGGCCCGCAATCTACG 24
Db      267 CAGCGCGAGGCCCGCAATCTACG 289

RESULT 18
US-09-489-039A-5991/c
; Sequence 5991, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5991
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5991

```

```
Query Match          63.8%; Score 16.6; DB 4; Length 1305;
Best Local Similarity 82.6%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGGAATCTAC 23
    |||||
Db 1141 ACATCGGCGAGCGCCGGAATGTAC 1119

RESULT 19
US-09-252-991A-1664
; Sequence 1664, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1664
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1664

Query Match          63.8%; Score 16.6; DB 4; Length 1350;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTACG 24
    |||||
Db 130 CAACGGCGAGCGCGGAGTCTATG 152

RESULT 20
US-09-252-991A-1456/c
; Sequence 1456, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1456
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1456

Query Match          63.8%; Score 16.6; DB 4; Length 1494;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTACG 24
    |||||
Db 682 CAACGGCGAGCGCGGAGTCTATG 660

RESULT 21
US-09-252-991A-1567
```

```
; Sequence 1567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1567
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1567

Query Match          63.8%; Score 16.6; DB 4; Length 1536;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTACG 24
    |||||
Db 855 CAACGGCGAGCGCGGAGTCTATG 877

RESULT 22
US-09-252-991A-15918/c
; Sequence 15918, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15918
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15918

Query Match          63.1%; Score 16.4; DB 4; Length 438;
Best Local Similarity 76.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGGAATCTACGAA 26
    |||||
Db 155 ACACGGCGAGCGCCGCGCAGCCAGGAA 130

RESULT 23
US-09-429-906B-1
; Sequence 1, Application US/09429906B
; Patent No. 6553099
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Orozco, Jr., Emil M.
; TITLE OF INVENTION: Plant UDP-Glucose Dehydrogenase
; FILE REFERENCE: BB1271 US NA
; CURRENT APPLICATION NUMBER: US/09/429,906B
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/107,274
; PRIOR FILING DATE: 1998-11-05
```



```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12299
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12299

Query Match      62.3%; Score 16.2; DB 4; Length 3063;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGAGGCCGCAATCTACGA 25
Db 1264 CGCGAGGCCGCACTACGA 1284

RESULT 28
US-09-252-991A-12350
; Sequence 1, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12350
; LENGTH: 3105
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12350

Query Match      62.3%; Score 16.2; DB 4; Length 3105;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGAGGCCGCAATCTACGA 25
Db 2007 CGCGAGGCCGCACTACGA 2027

RESULT 29
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-103-840A-1

Query Match      62.3%; Score 16.2; DB 3; Length 4411529;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGCAATCTA 22
Db 3575487 CAACGGCGTCCGCCGCAATCGA 3575467

Search completed: June 20, 2004, 11:46:40
Job time : 42.8763 secs
```

```
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      62.3%; Score 16.2; DB 3; Length 4403765;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGCAATCTA 22
Db 3570341 CAACGGCGTCCGCCGCAATCGA 3570321

RESULT 30
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-103-840A-1

Query Match      62.3%; Score 16.2; DB 3; Length 4411529;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGCAATCTA 22
Db 3575487 CAACGGCGTCCGCCGCAATCGA 3575467

Search completed: June 20, 2004, 11:46:40
Job time : 42.8763 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 147.058 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26
Sequence: 1 acaacgcgagggccgaatctacga 26

Scoring table: IDENTITY NUC
Gapop 10'0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| C 1 | 26 | 100.0 | 401 | 9 | US-09-956-634-75 |
| C 2 | 26 | 100.0 | 401 | 10 | US-09-997-182-75 |
| C 3 | 26 | 100.0 | 401 | 10 | US-09-997-181-75 |
| 4 | 18.6 | 71.5 | 505 | 13 | US-10-194-163-376 |
| 5 | 18 | 69.2 | 6357 | 13 | US-10-424-599-133268 |
| 6 | 17.8 | 68.5 | 806 | 13 | US-10-424-599-67461 |
| 7 | 17.6 | 67.7 | 531 | 15 | US-10-029-386-6930 |
| 8 | 17.6 | 67.7 | 612 | 13 | US-10-425-114-22881 |
| 9 | 17.6 | 67.7 | 1017 | 13 | US-10-282-122A-12032 |
| 10 | 17.6 | 67.7 | 2205 | 16 | US-10-369-493-44269 |
| C 11 | 17.6 | 67.7 | 536165 | 10 | US-09-939-964-1 |
| 12 | 17 | 65.4 | 383 | 13 | US-10-424-599-110822 |
| 13 | 17 | 65.4 | 669 | 15 | US-10-156-761-2378 |
| 14 | 17 | 65.4 | 672 | 16 | US-10-369-493-27924 |
| | | | | | Sequence 27924, A |
| | | | | | Sequence 75, Appl |
| | | | | | Sequence 75, Appl |
| | | | | | Sequence 376, Appl |
| | | | | | Sequence 132268, |
| | | | | | Sequence 67461, A |
| | | | | | Sequence 6930, Ap |
| | | | | | Sequence 22881, A |
| | | | | | Sequence 12032, A |
| | | | | | Sequence 44269, A |
| | | | | | Sequence 1, Appl |
| | | | | | Sequence 110822, |
| | | | | | Sequence 2378, Ap |
| | | | | | Sequence 27924, A |

| | | | | | | |
|------|----|------|---------|----|----------------------|--------------------|
| 65.4 | 17 | C 15 | 821 | 13 | US-10-027-632-168321 | Sequence 168321, |
| 65.4 | 17 | C 16 | 821 | 13 | US-10-027-632-168322 | Sequence 168322, |
| 65.4 | 17 | C 17 | 821 | 16 | US-10-027-632-168321 | Sequence 168321, |
| 65.4 | 17 | C 18 | 821 | 16 | US-10-027-632-168322 | Sequence 168322, |
| 65.4 | 17 | C 19 | 879 | 14 | US-10-027-806-37 | Sequence 37, Appl |
| 65.4 | 17 | C 20 | 879 | 14 | US-10-034-823-37 | Sequence 37, Appl |
| 65.4 | 17 | C 21 | 879 | 15 | US-10-027-801-37 | Sequence 37, Appl |
| 65.4 | 17 | C 22 | 879 | 15 | US-10-029-120-37 | Sequence 37, Appl |
| 65.4 | 17 | C 23 | 1173 | 13 | US-10-282-122A-13773 | Sequence 13773, A |
| 65.4 | 17 | C 24 | 1391 | 13 | US-10-425-114-23453 | Sequence 23453, A |
| 65.4 | 17 | C 25 | 1557 | 16 | US-10-369-493-28653 | Sequence 28653, A |
| 65.4 | 17 | C 26 | 1563 | 16 | US-10-369-493-31411 | Sequence 31411, A |
| 65.4 | 17 | C 27 | 1612 | 13 | US-10-425-114-260 | Sequence 260, App |
| 65.4 | 17 | C 28 | 1935 | 15 | US-10-156-761-5202 | Sequence 5202, App |
| 65.4 | 17 | C 29 | 42432 | 14 | US-10-027-806-2 | Sequence 2, Appl |
| 65.4 | 17 | C 30 | 42432 | 14 | US-10-034-623-2 | Sequence 2, Appl |
| 65.4 | 17 | C 31 | 42432 | 15 | US-10-027-801-2 | Sequence 2, Appl |
| 65.4 | 17 | C 32 | 42432 | 15 | US-10-029-120-2 | Sequence 2, Appl |
| 65.4 | 17 | C 33 | 93483 | 16 | US-10-034-650-49 | Sequence 49, Appl |
| 65.4 | 17 | C 34 | 9025608 | 15 | US-10-156-761-1 | Sequence 1, Appl |
| 65.4 | 17 | C 35 | 9025608 | 15 | US-10-156-761-1 | Sequence 1, Appl |
| 64.6 | 16 | C 36 | 1290 | 16 | US-10-369-493-44453 | Sequence 44453, A |
| 64.6 | 16 | C 37 | 1755 | 16 | US-10-260-238-721 | Sequence 721, App |
| 64.6 | 16 | C 38 | 5033 | 13 | US-10-453-468-1 | Sequence 1, Appl |
| 63.8 | 16 | C 39 | 290 | 9 | US-09-294-093B-3144 | Sequence 3144, App |
| 63.8 | 16 | C 40 | 864 | 16 | US-10-260-238-3741 | Sequence 3741, App |
| 63.8 | 16 | C 41 | 1284 | 16 | US-10-369-493-35911 | Sequence 35911, A |
| 63.8 | 16 | C 42 | 1407 | 16 | US-10-438-784-4 | Sequence 4, Appl |
| 63.8 | 16 | C 43 | 1455 | 13 | US-10-282-122A-17996 | Sequence 17996, A |
| 63.8 | 16 | C 44 | 2148 | 13 | US-10-425-114-3454 | Sequence 3454, App |
| 63.8 | 16 | C 45 | 2154 | 16 | US-10-369-493-39263 | Sequence 39263, A |
| 63.8 | 16 | C 46 | 2157 | 13 | US-10-282-122A-15107 | Sequence 15107, A |
| 63.8 | 16 | C 47 | 2160 | 16 | US-10-369-493-38639 | Sequence 38639, A |
| 63.8 | 16 | C 48 | 2160 | 16 | US-10-369-493-40003 | Sequence 40003, A |
| 63.8 | 16 | C 49 | 2466 | 16 | US-10-369-493-42318 | Sequence 42318, A |
| 63.8 | 16 | C 50 | 3807 | 16 | US-10-438-784-2 | Sequence 2, Appl |
| 63.8 | 16 | C 51 | 4218 | 15 | US-10-017-161-1969 | Sequence 1969, App |
| 63.8 | 16 | C 52 | 4218 | 16 | US-10-292-798-1617 | Sequence 1617, App |
| 63.8 | 16 | C 53 | 5898 | 16 | US-10-438-784-1 | Sequence 1, Appl |
| 63.8 | 16 | C 54 | 6077 | 15 | US-10-311-455-1828 | Sequence 1828, App |
| 63.8 | 16 | C 55 | 7543 | 10 | US-09-993-731-3 | Sequence 3, Appl |
| 63.8 | 16 | C 56 | 34769 | 13 | US-10-221-714A-502 | Sequence 502, App |
| 63.8 | 16 | C 57 | 67832 | 12 | US-09-997-722-67 | Sequence 67, Appl |
| 63.1 | 16 | C 58 | 266 | 13 | US-10-424-599-3052 | Sequence 3052, App |
| 63.1 | 16 | C 59 | 381 | 13 | US-10-424-599-142107 | Sequence 142107, |
| 63.1 | 16 | C 60 | 546 | 13 | US-10-227-964C-1 | Sequence 1, Appl |
| 63.1 | 16 | C 61 | 582 | 15 | US-10-156-761-6486 | Sequence 6486, App |
| 63.1 | 16 | C 62 | 627 | 16 | US-10-369-493-41107 | Sequence 41107, A |
| 63.1 | 16 | C 63 | 699 | 13 | US-10-282-122A-30637 | Sequence 30637, A |
| 63.1 | 16 | C 64 | 1422 | 13 | US-10-282-122A-25899 | Sequence 25899, A |
| 63.1 | 16 | C 65 | 1923 | 16 | US-10-369-493-44696 | Sequence 44696, A |
| 63.1 | 16 | C 66 | 2000 | 9 | US-09-938-842A-4884 | Sequence 4884, App |
| 63.1 | 16 | C 67 | 2000 | 11 | US-09-938-842A-4884 | Sequence 4884, App |
| 63.1 | 16 | C 68 | 2265 | 9 | US-09-738-626-435 | Sequence 435, App |
| 63.1 | 16 | C 69 | 2532 | 15 | US-10-156-761-100 | Sequence 100, App |
| 63.1 | 16 | C 70 | 2670 | 15 | US-10-128-714-153 | Sequence 153, App |
| 63.1 | 16 | C 71 | 2837 | 15 | US-10-128-714-5153 | Sequence 5153, App |
| 63.1 | 16 | C 72 | 6249 | 15 | US-10-311-455-1188 | Sequence 1188, App |
| 63.1 | 16 | C 73 | 17967 | 15 | US-10-311-455-987 | Sequence 987, App |
| 63.1 | 16 | C 74 | 46404 | 13 | US-10-087-132-634 | Sequence 634, App |
| 63.1 | 16 | C 75 | 330400 | 9 | US-08-738-626-1 | Sequence 1, Appl |
| 62.3 | 16 | C 76 | 214 | 13 | US-10-085-733A-44852 | Sequence 44852, A |
| 62.3 | 16 | C 77 | 214 | 16 | US-10-242-535A-44852 | Sequence 44852, A |
| 62.3 | 16 | C 78 | 330 | 11 | US-09-864-408A-5843 | Sequence 5843, App |
| 62.3 | 16 | C 79 | 471 | 9 | US-09-738-626-2454 | Sequence 2454, App |
| 62.3 | 16 | C 80 | 1185 | 15 | US-10-156-761-6562 | Sequence 6562, App |
| 62.3 | 16 | C 81 | 1353 | 13 | US-10-282-122A-14430 | Sequence 14430, A |
| 62.3 | 16 | C 82 | 1342 | 13 | US-10-282-122A-11876 | Sequence 11876, A |
| 62.3 | 16 | C 83 | 2073 | 13 | US-10-282-122A-26442 | Sequence 26442, A |
| 62.3 | 16 | C 84 | 2076 | 13 | US-10-282-122A-28167 | Sequence 28167, A |
| 62.3 | 16 | C 85 | 2748 | 9 | US-09-738-626-2853 | Sequence 2853, App |
| 62.3 | 16 | C 86 | 7256 | 17 | US-10-239-090A-44 | Sequence 44, Appl |
| 62.3 | 16 | C 87 | 8161 | 9 | US-09-759-152-9 | Sequence 9, Appl |

| | | | | | | | | | | | | | |
|-------|------|------|---------|----|----------------------|----------------------|-----|------|------|-------|----|----------------------|---------------------|
| 88 | 16.2 | 62.3 | 8175 | 9 | US-09-759-152-7 | Sequence 7, Appli | 161 | 15.6 | 60.0 | 460 | 13 | US-10-085-783A-58370 | Sequence 58370, A |
| c 89 | 16.2 | 62.3 | 150223 | 17 | US-10-624-149A-1 | Sequence 1, Appli | 162 | 15.6 | 60.0 | 460 | 16 | US-10-242-535A-58370 | Sequence 58370, A |
| 90 | 16.2 | 62.3 | 2940917 | 13 | US-10-027-632-174763 | Sequence 174763, A | 163 | 15.6 | 60.0 | 500 | 13 | US-10-085-783A-31184 | Sequence 31184, A |
| 91 | 16.2 | 62.3 | 2940917 | 16 | US-10-027-632-174763 | Sequence 174763, A | 164 | 15.6 | 60.0 | 500 | 16 | US-10-242-535A-31184 | Sequence 31184, A |
| 92 | 16.2 | 62.3 | 3309400 | 9 | US-09-738-626-1 | Sequence 1, Appli | 165 | 15.6 | 60.0 | 531 | 15 | US-10-156-761-5870 | Sequence 5870, A |
| 93 | 16 | 61.5 | 286 | 11 | US-09-864-408A-3873 | Sequence 3873, App | 166 | 15.6 | 60.0 | 546 | 9 | US-09-864-761-13268 | Sequence 13268, A |
| 94 | 16 | 61.5 | 363 | 15 | US-10-156-761-721 | Sequence 721, App | 167 | 15.6 | 60.0 | 574 | 9 | US-09-864-761-13696 | Sequence 13696, A |
| 95 | 16 | 61.5 | 385 | 13 | US-10-424-599-134331 | Sequence 134331, App | 168 | 15.6 | 60.0 | 629 | 13 | US-10-027-632-285790 | Sequence 285790, A |
| 96 | 16 | 61.5 | 387 | 9 | US-09-867-701-1569 | Sequence 1569, App | 169 | 15.6 | 60.0 | 629 | 16 | US-10-027-632-285790 | Sequence 285790, A |
| 97 | 16 | 61.5 | 406 | 10 | US-09-803-719-2368 | Sequence 2368, App | 170 | 15.6 | 60.0 | 756 | 16 | US-10-369-493-43105 | Sequence 43105, A |
| 98 | 16 | 61.5 | 441 | 15 | US-10-156-761-1893 | Sequence 1893, App | 171 | 15.6 | 60.0 | 771 | 16 | US-10-369-493-33292 | Sequence 33292, A |
| 99 | 16 | 61.5 | 489 | 9 | US-09-894-844-63 | Sequence 63, Appli | 172 | 15.6 | 60.0 | 830 | 13 | US-10-282-122A-29974 | Sequence 29974, A |
| 100 | 16 | 61.5 | 489 | 13 | US-10-647-089-63 | Sequence 63, Appli | 173 | 15.6 | 60.0 | 832 | 13 | US-10-221-625-207 | Sequence 207, App |
| 101 | 16 | 61.5 | 489 | 16 | US-10-388-302-63 | Sequence 63, Appli | 174 | 15.6 | 60.0 | 870 | 15 | US-10-156-761-3286 | Sequence 3286, App |
| 102 | 16 | 61.5 | 498 | 16 | US-10-260-238-3297 | Sequence 3297, App | 175 | 15.6 | 60.0 | 872 | 16 | US-10-260-238-3766 | Sequence 3766, App |
| 103 | 16 | 61.5 | 535 | 13 | US-10-424-599-131246 | Sequence 131246, A | 176 | 15.6 | 60.0 | 872 | 16 | US-10-156-761-1469 | Sequence 1469, App |
| 104 | 16 | 61.5 | 549 | 13 | US-10-621-901-201 | Sequence 201, App | 177 | 15.6 | 60.0 | 1053 | 15 | US-10-240-965-56 | Sequence 56, Appli |
| c 105 | 16 | 61.5 | 604 | 13 | US-10-027-632-275432 | Sequence 275432, A | 178 | 15.6 | 60.0 | 1075 | 15 | US-10-219-220-216 | Sequence 216, App |
| c 106 | 16 | 61.5 | 604 | 16 | US-10-027-632-275432 | Sequence 275432, A | 179 | 15.6 | 60.0 | 1125 | 15 | US-10-156-761-2293 | Sequence 2293, App |
| 107 | 16 | 61.5 | 779 | 16 | US-10-027-632-154384 | Sequence 154384, A | 180 | 15.6 | 60.0 | 1131 | 16 | US-10-369-493-33513 | Sequence 33513, A |
| 108 | 16 | 61.5 | 779 | 16 | US-10-027-632-154384 | Sequence 154384, A | 181 | 15.6 | 60.0 | 1287 | 16 | US-10-260-238-845 | Sequence 845, App |
| 109 | 16 | 61.5 | 941 | 10 | US-09-991-936-1597 | Sequence 1597, App | 182 | 15.6 | 60.0 | 1389 | 15 | US-10-230-026-27 | Sequence 27, Appli |
| 110 | 16 | 61.5 | 969 | 13 | US-10-282-122A-13043 | Sequence 13043, A | 183 | 15.6 | 60.0 | 1413 | 16 | US-10-369-493-35527 | Sequence 35527, A |
| 111 | 16 | 61.5 | 969 | 13 | US-10-282-122A-14234 | Sequence 14234, A | 184 | 15.6 | 60.0 | 1443 | 15 | US-10-156-761-3397 | Sequence 3397, App |
| 112 | 16 | 61.5 | 971 | 9 | US-09-879-957-197 | Sequence 197, App | 185 | 15.6 | 60.0 | 1458 | 16 | US-10-369-493-42392 | Sequence 42392, A |
| 113 | 16 | 61.5 | 988 | 16 | US-10-198-706-1 | Sequence 1, Appli | 186 | 15.6 | 60.0 | 1470 | 13 | US-10-282-122A-25366 | Sequence 25366, A |
| c 114 | 16 | 61.5 | 1066 | 16 | US-10-260-238-4037 | Sequence 4037, App | 187 | 15.6 | 60.0 | 1508 | 16 | US-10-369-493-41474 | Sequence 41474, A |
| c 115 | 16 | 61.5 | 1080 | 13 | US-10-282-122A-33483 | Sequence 33483, A | 188 | 15.6 | 60.0 | 1508 | 16 | US-10-282-122A-25889 | Sequence 25889, A |
| 116 | 16 | 61.5 | 1080 | 15 | US-10-156-761-4054 | Sequence 4054, App | 189 | 15.6 | 60.0 | 2139 | 13 | US-10-282-122A-25889 | Sequence 25889, A |
| 117 | 16 | 61.5 | 1116 | 15 | US-10-156-761-6640 | Sequence 6640, App | 190 | 15.6 | 60.0 | 2160 | 13 | US-10-430-011-76 | Sequence 76, Appli |
| 118 | 16 | 61.5 | 1135 | 13 | US-10-424-599-33781 | Sequence 33781, A | 191 | 15.6 | 60.0 | 2160 | 13 | US-10-430-011-83 | Sequence 83, Appli |
| 119 | 16 | 61.5 | 1167 | 16 | US-10-369-493-34189 | Sequence 34189, A | 192 | 15.6 | 60.0 | 2160 | 13 | US-10-430-011-126 | Sequence 126, App |
| 120 | 16 | 61.5 | 1188 | 15 | US-10-156-761-6223 | Sequence 6223, App | 193 | 15.6 | 60.0 | 2160 | 15 | US-10-138-927-76 | Sequence 76, Appli |
| 121 | 16 | 61.5 | 1197 | 15 | US-10-156-761-5205 | Sequence 5205, App | 194 | 15.6 | 60.0 | 2160 | 15 | US-10-138-927-83 | Sequence 83, Appli |
| c 122 | 16 | 61.5 | 1227 | 9 | US-09-974-300-1174 | Sequence 1174, App | 195 | 15.6 | 60.0 | 2291 | 13 | US-10-282-122A-14324 | Sequence 14324, A |
| 123 | 16 | 61.5 | 1242 | 10 | US-09-910-186A-21 | Sequence 21, Appli | 196 | 15.6 | 60.0 | 2337 | 13 | US-10-282-122A-25485 | Sequence 25485, A |
| 124 | 16 | 61.5 | 1302 | 15 | US-10-156-761-4554 | Sequence 4554, App | 197 | 15.6 | 60.0 | 2455 | 16 | US-10-369-493-36725 | Sequence 36725, A |
| 125 | 16 | 61.5 | 1534 | 13 | US-10-087-192-143 | Sequence 143, App | 198 | 15.6 | 60.0 | 2517 | 13 | US-10-282-122A-25466 | Sequence 25466, A |
| 126 | 16 | 61.5 | 1557 | 16 | US-10-369-493-44159 | Sequence 44159, A | 199 | 15.6 | 60.0 | 2547 | 13 | US-10-282-122A-26047 | Sequence 26047, A |
| 127 | 16 | 61.5 | 1737 | 17 | US-10-276-751-2 | Sequence 2, Appli | 200 | 15.6 | 60.0 | 2547 | 16 | US-10-369-493-24476 | Sequence 24476, A |
| 128 | 16 | 61.5 | 1737 | 17 | US-10-276-751-3 | Sequence 3, Appli | 201 | 15.6 | 60.0 | 2547 | 13 | US-10-412-599B-1584 | Sequence 1584, App |
| 129 | 16 | 61.5 | 1872 | 13 | US-10-425-114-15187 | Sequence 15187, A | 202 | 15.6 | 60.0 | 2849 | 15 | US-10-156-761-361 | Sequence 361, App |
| 130 | 16 | 61.5 | 1881 | 16 | US-10-369-493-28460 | Sequence 28460, A | 203 | 15.6 | 60.0 | 3750 | 13 | US-10-282-122A-27453 | Sequence 27453, A |
| 131 | 16 | 61.5 | 1881 | 16 | US-10-369-493-31420 | Sequence 31420, A | 204 | 15.6 | 60.0 | 10396 | 13 | US-10-282-122A-33665 | Sequence 33665, A |
| 132 | 16 | 61.5 | 1884 | 17 | US-10-276-751-1 | Sequence 1, Appli | 205 | 15.6 | 60.0 | 11050 | 15 | US-10-204-708-85 | Sequence 85, Appli |
| 133 | 16 | 61.5 | 1932 | 16 | US-10-369-493-41722 | Sequence 41722, A | 206 | 15.6 | 60.0 | 11275 | 15 | US-10-240-689-39 | Sequence 39, Appli |
| 134 | 16 | 61.5 | 2034 | 15 | US-10-156-761-3529 | Sequence 3529, App | 207 | 15.4 | 59.2 | 237 | 9 | US-09-864-761-28053 | Sequence 28053, A |
| 135 | 16 | 61.5 | 2055 | 13 | US-10-424-599-79230 | Sequence 79230, A | 208 | 15.4 | 59.2 | 253 | 13 | US-10-424-599-51467 | Sequence 51467, A |
| 136 | 16 | 61.5 | 2205 | 13 | US-10-424-599-79230 | Sequence 79230, A | 209 | 15.4 | 59.2 | 271 | 13 | US-10-424-599-86539 | Sequence 86539, A |
| 137 | 16 | 61.5 | 2364 | 13 | US-10-282-122A-11362 | Sequence 11362, A | 210 | 15.4 | 59.2 | 275 | 9 | US-09-923-876-278 | Sequence 278, App |
| 138 | 16 | 61.5 | 2559 | 14 | US-10-011-588-24 | Sequence 24, Appli | 211 | 15.4 | 59.2 | 275 | 13 | US-10-424-599-10860 | Sequence 10860, A |
| c 139 | 16 | 61.5 | 2686 | 13 | US-10-424-599-33779 | Sequence 33779, A | 212 | 15.4 | 59.2 | 293 | 13 | US-10-029-386-10860 | Sequence 10860, A |
| 140 | 16 | 61.5 | 3318 | 13 | US-10-282-122A-31347 | Sequence 31347, A | 213 | 15.4 | 59.2 | 327 | 15 | US-10-029-386-10860 | Sequence 10860, A |
| 141 | 16 | 61.5 | 4079 | 16 | US-10-305-720-1248 | Sequence 1248, App | 214 | 15.4 | 59.2 | 369 | 13 | US-10-424-599-63343 | Sequence 63343, A |
| 142 | 16 | 61.5 | 4519 | 10 | US-09-360-706-712 | Sequence 712, App | 215 | 15.4 | 59.2 | 443 | 13 | US-10-027-632-289663 | Sequence 289663, A |
| 143 | 16 | 61.5 | 4519 | 17 | US-10-641-643-1202 | Sequence 1202, App | 216 | 15.4 | 59.2 | 443 | 16 | US-10-027-632-289663 | Sequence 289663, A |
| 144 | 16 | 61.5 | 4519 | 17 | US-10-283-975A-223 | Sequence 223, App | 217 | 15.4 | 59.2 | 447 | 13 | US-10-425-114-17731 | Sequence 17731, A |
| 145 | 16 | 61.5 | 4851 | 13 | US-10-282-122A-25489 | Sequence 25489, A | 218 | 15.4 | 59.2 | 462 | 9 | US-09-864-761-11462 | Sequence 11462, A |
| 146 | 16 | 61.5 | 88421 | 9 | US-09-976-059-1 | Sequence 1, Appli | 219 | 15.4 | 59.2 | 463 | 13 | US-10-424-599-125650 | Sequence 125650, A |
| 147 | 16 | 61.5 | 169276 | 13 | US-10-087-192-142 | Sequence 142, App | 220 | 15.4 | 59.2 | 463 | 13 | US-10-282-122A-23115 | Sequence 23115, A |
| 148 | 16 | 61.5 | 3186778 | 13 | US-10-027-632-174961 | Sequence 174961, A | 221 | 15.4 | 59.2 | 477 | 13 | US-10-424-599-76495 | Sequence 76495, A |
| 149 | 16 | 61.5 | 3186778 | 16 | US-10-027-632-174961 | Sequence 174961, A | 222 | 15.4 | 59.2 | 480 | 15 | US-10-156-761-3954 | Sequence 3954, App |
| 150 | 15.8 | 60.8 | 444 | 15 | US-10-156-761-289 | Sequence 289, App | 223 | 15.4 | 59.2 | 488 | 16 | US-10-260-238-111 | Sequence 111, App |
| 151 | 15.8 | 60.8 | 1119 | 16 | US-10-369-493-28235 | Sequence 28235, A | 224 | 15.4 | 59.2 | 507 | 13 | US-10-424-599-65821 | Sequence 65821, A |
| 152 | 15.8 | 60.8 | 1206 | 13 | US-10-282-122A-13454 | Sequence 13454, A | 225 | 15.4 | 59.2 | 507 | 15 | US-10-156-761-2692 | Sequence 2692, App |
| 153 | 15.8 | 60.8 | 2529 | 15 | US-10-156-761-6449 | Sequence 6449, App | 226 | 15.4 | 59.2 | 520 | 13 | US-10-424-599-125334 | Sequence 125334, A |
| c 154 | 15.8 | 60.8 | 3865 | 13 | US-10-087-192-964 | Sequence 964, App | 227 | 15.4 | 59.2 | 550 | 15 | US-10-106-698-2652 | Sequence 2652, App |
| c 155 | 15.6 | 60.0 | 334 | 9 | US-09-864-761-29834 | Sequence 29834, A | 228 | 15.4 | 59.2 | 551 | 15 | US-10-029-386-12136 | Sequence 12136, App |
| c 156 | 15.6 | 60.0 | 333 | 16 | US-10-085-783A-13248 | Sequence 13248, A | 229 | 15.4 | 59.2 | 571 | 10 | US-09-918-995-6136 | Sequence 6136, App |
| 157 | 15.6 | 60.0 | 333 | 16 | US-10-242-535A-13248 | Sequence 13248, A | 230 | 15.4 | 59.2 | 579 | 13 | US-10-027-632-106025 | Sequence 106025, A |
| c 158 | 15.6 | 60.0 | 382 | 9 | US-09-783-590-6427 | Sequence 6427, App | 231 | 15.4 | 59.2 | 579 | 13 | US-10-027-632-106026 | Sequence 106026, A |
| c 159 | 15.6 | 60.0 | 441 | 15 | US-10-156-761-5497 | Sequence 5497, App | 232 | 15.4 | 59.2 | 579 | 16 | US-10-027-632-106025 | Sequence 106025, A |
| 160 | 15.6 | 60.0 | 456 | 15 | US-10-219-220-14 | Sequence 14, Appli | 233 | 15.4 | 59.2 | 579 | 16 | US-10-027-632-106026 | Sequence 106026, A |

RESULT 1

```
Query Match      100.0%; Score 26; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 0.0076;
Matches 26; Conservative 0; Mismatches 0; Indels
```

Qy 1 ACAACGGCGAGCCCGAATCTACGAA 26
db 148 ACAACGGCGAGCCCGAATCTACGAA 123

RESULT 3

```

US-09-997-181-75/c
; Sequence 75, Application US/09997181
; Publication NO. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory peptides
; FILE REFERENCE: 61257
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 75
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(401)
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.
; 09-997-181-75

```

Query Match 100.0%; Score 26; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 0.0076;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCGAATCTACGAA 26
|||
Db 148 ACAACGGCGAGGCCGAATCTACGAA 123

RESULT 4

US-10-194-163-376
; Sequence 376, Application US/10194163
; Publication No. US20020172976A1
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce Carter
; TITLE OF INVENTION: PORPHYROMAS GINGIVALIS POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Fast-SEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/194,163
; FILING DATE: 04-Nov-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Basu, Shantanu

REGISTRATION NUMBER: 43,318
REFERENCE/DOCKET NUMBER: 529282000101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5995
TELEFAX: 650-494-0792
TELEX: 706141

TELEX: 706141
 INFORMATION FOR SEQ ID NO: 376
 SEQUENCE CHARACTERISTICS:
 LENGTH: 505 base pairs
 TYPE: nucleic acid
 TOPOLOGY: double
 STRANDEDNESS: double
 MOLECULE TYPE: circular
 HYPOTHETICAL: NO
 ANTI-SENSE: UNKNOWN
 ORIGINAL SOURCE:
 ORGANISM: PORPHYROMONAS GINGIVALIS
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1..505
 SEQUENCE DESCRIPTION: SEQ ID NO: 376
 US-10-194-163-376

| | | | | |
|-----------------------|-----------------|---------------|-----------|-------------|
| Query Match | 71.5%; | Score 18.6; | DB 13; | Length 505; |
| Best Local Similarity | 84.0%; | Pred. No. 29; | | |
| Matches 21; | Conservative 0; | Mismatches 4; | Indels 0; | Gaps 0; |

QY 2 CAACGGCGAGGCCGAATCTACGA 26

Db 336 CAACGGCATGGCCGGAATCAACGAA 360

RESULTS

```

US-10-424-599-132268
/ / Sequence 132268, Application US/10424599
/ / Publication NO. US20040031072A1
/ / GENERAL INFORMATION:
/ / APPLICANT: La Rosa Thomas J
/ / APPLICANT: Kovalic David K
/ / APPLICANT: Zhou Yihua
/ / APPLICANT: Cao Yongwei
/ / TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ / TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ / FILE REFERENCE: 38-21(53223) B
/ / CURRENT APPLICATION NUMBER: US/10/424,599
/ / CURRENT FILING DATE: 2003-04-28
/ / NUMBER OF SEQ ID NOS: 285684
/ / SEQ ID NO 132268
/ / LENGTH: 6357
/ / TYPE: DNA
/ / ORGANISM: Glycine max
/ / FEATURE:
/ / OTHER INFORMATION: Clone ID: PAT_MRT3847_90445C.1
US-10-424-599-132268

```

```
Query Match      69.2%; Score 18; DB 13; Length 6357;
Best Local Similarity 80.8%; Pred. No. 55;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

QY . 1 ACAACGGCGAGGCCCGAATCTACGAA 26
D**b** 401 ACGACGTCGAGGCCCTGATCGACGAA 426

RESIST. T. 6

US-10-424-598-67461
; Sequence 67461, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 67461
; LENGTH: 806
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_31931C.1
US-10-424-599-67461

Query Match      58.5%; Score 17.8; DB 13; Length 806;
Best Local Similarity 90.5%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5 CGGCGAGGCCCGAATCTACGA 25
Db      444 CGGCGAGGCCCGGATCAACGA 464

RESULT 7
US-10-029-386-6930
; Sequence 6930, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6930
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009336.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q03828, EVALUE 3.00e-18
; OTHER INFORMATION: EST HUMAN HIT: BG489979.1, EVALUE 4.80e-02
; OTHER INFORMATION: NT HIT: S34322.1, EVALUE 1.00e-41
US-10-029-386-6930

Query Match      67.7%; Score 17.6; DB 15; Length 531;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 AACGGCGAGGCCCGAATCTACGAA 26
Db      506 AACGGCGTGGCCCGAGCAACGGA 529

RESULT 8
US-10-425-114-22681
; Sequence 22681, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
```

```
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22681
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-095-F7_FLI
US-10-425-114-22681

Query Match      67.7%; Score 17.6; DB 13; Length 612;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CAACGGCGAGGCCCGAATCTACGA 25
Db      174 CAACGGCGGCGCTACATCTACGA 197

RESULT 9
US-10-282-122A-12032
; Sequence 12032, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA_034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12032
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
```

US-10-282-122A-12032

Query Match 67.7%; Score 17.6; DB 13; Length 1017;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGCAATCTACGA 25

Db 798 CATCGCGAGCCCGCACTACGA 821

RESULT 10

US-10-369-493-44269
; Sequence 44269, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44269
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44269

Query Match 67.7%; Score 17.6; DB 16; Length 2205;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGCCCGCAATCTACGA 26

Db 1996 AACGGCGAGTCGAGATCTACGA 2019

RESULT 11

US-09-939-964-1/c
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; FILE REFERENCE: Plasmid
; CURRENT APPLICATION NUMBER: US/09/939,964
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/214,808
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-939-964-1

Query Match 67.7%; Score 17.6; DB 10; Length 536165;
Best Local Similarity 83.3%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGCCCGCAATCTACGA 26

Db 85320 AAGGGCGGCGCCCGAACAACGAA 85297

RESULT 12

US-10-424-599-110822
; Sequence 110822, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 110822
; LENGTH: 383
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_71082C.1
US-10-424-599-110822

Query Match 65.4%; Score 17; DB 13; Length 383;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGCAATCTACGA 25

Db 315 ACAACAGCTAGGCGCTGAATTTCCGA 339

RESULT 13

US-10-156-761-2378
; Sequence 2378, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, WASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2378
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(669)
US-10-156-761-2378

Query Match 65.4%; Score 17; DB 15; Length 669;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGCAATCTACGA 25

Db 278 AGAACGTCGAGCCCGTATCTACGA 302

RESULT 14

US-10-369-493-27924

Sequence 27924, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 27924

LENGTH: 672

TYPE: DNA

ORGANISM: Neurospora crassa

US-10-369-493-27924

Query Match 55.4%; Score 17; DB 15; Length 672;

Best Local Similarity 80.0%; Pred. No. 1.7e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGGAATCTACGAA 26

||||| ||||| ||||| ||||| |||||

Db 303 CAACCCAGAGACCGCATCTACGAA 327

RESULT 15

US-10-027-632-168321/c

Sequence 168321, Application US/10027632

Publication No. US20020198371A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 168321

LENGTH: 821

TYPE: DNA

ORGANISM: Human

US-10-027-632-168321

Query Match 55.4%; Score 17; DB 13; Length 821;

Best Local Similarity 80.0%; Pred. No. 1.7e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGGAATCTACGAA 26

```

; SEQ ID NO 168321
; LENGTH: 821
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-168321

```

Query Match 65.4%; Score 17; DB 16; Length 821;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels

QY 2 CAACGGCGAGGCCGAATCTACGAA 26
|||||
Dp 509 CAACCCCGAGGCATGAATCTACCAA 485
|||||

```

RESULT 18
US-10-027-632-168322/c
/ Sequence 168322, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108822.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO:168322
/ LENGTH: 821
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-168322

```

| | | | | |
|-----------------------|----------------|----------------------------|----------|------------|
| Query Match | 65.4% | Score 17 | DB 16 | Length 821 |
| Best Local Similarity | 80.0% | Pred. No. 1.7e+02 | | |
| Matches 20 | Conservative 0 | Mismatches 5 | Indels 0 | Gaps 0 |
| QY | 2 | CAACGGCGAGGCCCGCAATCTACGAA | 26 | |
| | | | | |
| Db | 509 | CAACCCCGAGGCATGCAATCTACCAA | 485 | |

```

RESULT 19
US-10-027-806-37/c
; Sequence 37, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feidman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP-002A
; CURRENT APPLICATION NUMBER: US/10/027, 806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 37
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(879)
US-16-027-806-37

```

```

Query Match      65.4%; Score 17; DB 14; Length 879;
Best Local Similarity 80.0%; Prid. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  ACAAGCGAGGCCCGCAATCTACGA 25
          434  ACCTCGAGAGGCCCGGAGTCTACAA 410

```

```

RESULT 20
US-10-034-623-37/c
; Sequence 37, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(879)
US-10-034-623-37

```

```

Query Match      65.4%; Score 17; DB 14; Length 879;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 ACAACGGCGAGGCCGCAATCTACGA 25
          |||||
DB      434 ACCTCGGAGAGGCCGAGTCTACAA 410

```

RESULT 21
US-10-027-801-37/c
/ Sequence 37, Application US/10027801
/ Publication No. US20030054364A1
/ GENERAL INFORMATION:
/ APPLICANT: Swanson, Ronald V.
/ APPLICANT: Feldman, Robert A.
/ APPLICANT: Schleper, Christa
/ TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSIUM
/ FILE REFERENCE: DCOEP.002A
/ CURRENT APPLICATION NUMBER: US/10/027,801
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
/ PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
/ NUMBER OF SEQ ID NOS: 123
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 37
/ LENGTH: 879
/ TYPE: DNA


```

; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(879)
US-10-027-801-37

Query Match      65.4%; Score 17; DB 15; Length 879;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 434 ACCTCGGAGAGGCCGCGAGTCTACAA 410

RESULT 22
US-10-029-120-37/c
; Sequence 37, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DORP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(879)
US-10-029-120-37

Query Match      65.4%; Score 17; DB 15; Length 879;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 434 ACCTCGGAGAGGCCGCGAGTCTACAA 410

RESULT 23
US-10-282-122A-13773/c
; Sequence 13773, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
;

; ORGANISM: Burkholderia fungorum
US-10-282-122A-13773

Query Match      65.4%; Score 17; DB 13; Length 1173;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGGCCGGAATCTACGAA 26
Db 199 CGACGGCGAGGCCGCGCTTCGAA 175

RESULT 24
US-10-425-114-23453/c
; Sequence 23453, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23453
; LENGTH: 1391
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3596-038-G7_FLI
US-10-425-114-23453

Query Match      65.4%; Score 17; DB 13; Length 1391;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 392 ACACGGCGGCGCTGCATCGAGA 368

RESULT 25
US-10-369-493-28653
; Sequence 28653, Application US/10369493
; Publication No. US20030233675A1
;
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13773
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13773
```


;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
;; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
;; NUMBER OF SEQ ID NOS: 123
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 42432
;; TYPE: DNA
;; ORGANISM: Cenarchaeum symbiosum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (3)...(10421)
;; NAME/KEY: CDS
;; LOCATION: (10625)...(11434)
;; NAME/KEY: CDS
;; LOCATION: (11478)...(13046)
;; NAME/KEY: CDS
;; LOCATION: (13046)...(14620)
;; NAME/KEY: CDS
;; LOCATION: (23558)...(24862)
;; NAME/KEY: CDS
;; LOCATION: (24913)...(25728)
;; NAME/KEY: CDS
;; LOCATION: (26504)...(26881)
;; NAME/KEY: CDS
;; LOCATION: (29655)...(30491)
;; NAME/KEY: CDS
;; LOCATION: (34559)...(36067)
;; NAME/KEY: CDS
;; LOCATION: (37002)...(37403)
;; NAME/KEY: CDS
;; LOCATION: (37404)...(38282)
;; NAME/KEY: CDS
;; LOCATION: (39454)...(40572)
;; US-10-027-806-2

Query Match 65.4%; Score 17; DB 14; Length 42432;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGCGCCGAATCTACGA 25
Db 37837 ACCTCGGAGAGCGCCGAGTCTACAA 37813

RESULT 30
US-10-034-623-2/c
;; Sequence 2, Application US/10034623
;; Publication No. US20020198365A1
;; GENERAL INFORMATION:
;; APPLICANT: Swanson, Ronald V.
;; APPLICANT: Feldman, Robert A.
;; APPLICANT: Schleper, Christa
;; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
;; FILE REFERENCE: DCOIP.002A
;; CURRENT APPLICATION NUMBER: US/10/034,623
;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: 09/408,020
;; PRIOR FILING DATE: 1999-09-29
;; PRIOR APPLICATION NUMBER: 60/102,294
;; PRIOR FILING DATE: 1998-09-29
;; NUMBER OF SEQ ID NOS: 123
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 42432
;; TYPE: DNA
;; ORGANISM: Cenarchaeum symbiosum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (3)...(10421)
;; NAME/KEY: CDS
;; LOCATION: (10625)...(11434)
;; NAME/KEY: CDS
;; LOCATION: (11478)...(13046)
;; NAME/KEY: CDS
;; LOCATION: (13046)...(14620)
;; NAME/KEY: CDS
;; LOCATION: (23558)...(24862)
;; NAME/KEY: CDS
;; LOCATION: (24913)...(25728)
;; NAME/KEY: CDS
;; LOCATION: (26504)...(26881)
;; NAME/KEY: CDS
;; LOCATION: (29655)...(30491)
;; NAME/KEY: CDS
;; LOCATION: (34559)...(36067)
;; NAME/KEY: CDS
;; LOCATION: (37002)...(37403)
;; NAME/KEY: CDS
;; LOCATION: (37404)...(38282)
;; NAME/KEY: CDS
;; LOCATION: (39454)...(40572)
;; US-10-027-806-2

;; LOCATION: (11478)...(13046)
;; NAME/KEY: CDS
;; LOCATION: (13046)...(14620)
;; NAME/KEY: CDS
;; LOCATION: (23558)...(24862)
;; NAME/KEY: CDS
;; LOCATION: (24913)...(25728)
;; NAME/KEY: CDS
;; LOCATION: (26504)...(26881)
;; NAME/KEY: CDS
;; LOCATION: (29655)...(30491)
;; NAME/KEY: CDS
;; LOCATION: (34559)...(36067)
;; NAME/KEY: CDS
;; LOCATION: (37002)...(37403)
;; NAME/KEY: CDS
;; LOCATION: (37404)...(38282)
;; NAME/KEY: CDS
;; LOCATION: (39454)...(40572)
;; US-10-034-623-2

Query Match 65.4%; Score 17; DB 14; Length 42432;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGCGCCGAATCTACGA 25
Db 37837 ACCTCGGAGAGCGCCGAGTCTACAA 37813

Search completed: June 20, 2004, 17:46:13
Job time : 162.058 secs

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|--------|--------|-----------|----|--------------------|
| | | Match | % | | | | |
| 1 | 19 | 100.0 | 393 | 2 | AA1749174 | | |
| C 2 | 19 | 100.0 | 110000 | 4 | AA199682 | 42 | Partial D |
| C 3 | 19 | 100.0 | 110000 | 4 | AA199683 | 42 | Continuation (43 o |
| C 4 | 17 | 89.5 | 25 | 8 | AA137369 | | Continuation (43 o |
| C 5 | 16.4 | 86.3 | 270 | 6 | AB172618 | | Ac137369 Human mic |
| C 6 | 16.4 | 86.3 | 432 | 3 | AA329698 | | Ab172618 Corn tass |
| C 7 | 16.4 | 86.3 | 445 | 3 | AA411047 | | AA329698 Zea may |
| C 8 | 16.4 | 86.3 | 963 | 7 | ADA70332 | | Aac411047 Zea may |
| C 9 | 16.4 | 86.3 | 1077 | 8 | ADA48367 | | Ada70332 Rice gene |
| C 10 | 16.4 | 86.3 | 1560 | 6 | AD229838 | | Ada48367 Rice gene |
| C 11 | 16.4 | 86.3 | 1348 | 7 | AC227231 | | AD229838 Rice NADP |
| 12 | 16.4 | 86.3 | 2000 | 6 | AB216341 | | AC227231 Prokaryot |
| 13 | 16.4 | 86.3 | 47475 | 3 | AAA81465 | 5 | Ab216341 Arabidops |
| 14 | 16.4 | 86.3 | 110000 | 3 | AAA81465 | | AAA81465 N. mening |
| 15 | 16.4 | 86.3 | 110000 | 4 | AA199682 | 02 | Continuation (6 of |
| 16 | 16.4 | 86.3 | 110000 | 4 | AA199683 | 02 | Continuation (3 of |
| 17 | 16.4 | 86.3 | 349980 | 3 | AA221612 | | Continuation (3 of |
| 18 | 16.4 | 86.3 | 349980 | 6 | AB281846 | | AA221612 Neisseria |
| C 19 | 16 | 84.2 | 942 | 9 | ADC23768 | | AB281846 Bifidobac |
| 20 | 16 | 84.2 | 1821 | 3 | AA453972 | | ADC23768 DNA sequ |
| 21 | 15.8 | 83.2 | 51 | 4 | AA175554 | | AA453972 M. tuberc |
| 22 | 15.8 | 83.2 | 51 | 4 | AA175555 | | AA175554 Human sil |
| 23 | 15.8 | 83.2 | 51 | 4 | AA179702 | | AA175555 Human sil |
| 24 | 15.8 | 83.2 | 51 | 4 | AA179702 | | AA179702 Human sil |

| | | | | | | | | | | | | | | | |
|-----|------|------|--------|---|-------------|--------------------|------------|------|------|------|-----|----------|----------|-----------|------------|
| 97 | 15.4 | 81.1 | 530 | 6 | ABQ17185 | Abq17185 | Oligonucle | 170 | 14.8 | 77.9 | 120 | 6 | ABL75421 | Ab175421 | Corn tass |
| 98 | 15.4 | 81.1 | 530 | 6 | ABQ17184 | Abq17184 | Oligonucle | 171 | 14.8 | 77.9 | 221 | 6 | ABN18064 | Abn18064 | Human ORF |
| 99 | 15.4 | 81.1 | 610 | 7 | ABX96466 | Abx96466 | Rice endo | 172 | 14.8 | 77.9 | 234 | 9 | ADD16092 | Add16092 | CDNA (Seq |
| 100 | 15.4 | 81.1 | 729 | 6 | ABQ26106 | Abq26106 | Oligonucle | 173 | 14.8 | 77.9 | 271 | 7 | ABX84396 | Abx84396 | Corn ear- |
| 101 | 15.4 | 81.1 | 729 | 6 | ABQ26107 | Abq26107 | Oligonucle | 174 | 14.8 | 77.9 | 282 | 7 | ABX87965 | Abx87965 | Corn ear- |
| 102 | 15.4 | 81.1 | 732 | 7 | ACA25151 | AcA25151 | Prokaryot | 175 | 14.8 | 77.9 | 298 | 7 | ABX87583 | Abx87583 | Corn ear- |
| 103 | 15.4 | 81.1 | 735 | 7 | ACA38066 | AcA38066 | Prokaryot | 176 | 14.8 | 77.9 | 315 | 3 | AAA31887 | Aaa31887 | Plant mic |
| 104 | 15.4 | 81.1 | 825 | 7 | ACA45245 | AcA45245 | Prokaryot | 177 | 14.8 | 77.9 | 318 | 6 | ABN76240 | Abn76240 | Human ORF |
| 105 | 15.4 | 81.1 | 851 | 6 | ABQ29083 | Abq29083 | Oligonucle | 178 | 14.8 | 77.9 | 354 | 8 | ACL11875 | AcL11875 | DNA clone |
| 106 | 15.4 | 81.1 | 851 | 6 | ABQ29082 | Abq29082 | Oligonucle | 179 | 14.8 | 77.9 | 355 | 2 | AAA34158 | Aaa34158 | Mycobacte |
| 107 | 15.4 | 81.1 | 903 | 6 | ABQ14805 | Abq14805 | Oligonucle | 180 | 14.8 | 77.9 | 356 | 2 | AAA34157 | Aaa34157 | Mycobacte |
| 108 | 15.4 | 81.1 | 903 | 6 | ABQ14804 | Abq14804 | Oligonucle | 181 | 14.8 | 77.9 | 357 | 2 | AAA34156 | Aaa34156 | Mycobacte |
| 109 | 15.4 | 81.1 | 945 | 7 | ACA35853 | AcA35853 | Prokaryot | 182 | 14.8 | 77.9 | 392 | 4 | AAK56036 | Aak56036 | Human inm |
| 110 | 15.4 | 81.1 | 981 | 7 | ACA49565 | AcA49565 | Prokaryot | 183 | 14.8 | 77.9 | 438 | 6 | ABN17464 | Abn17464 | Human ORF |
| 111 | 15.4 | 81.1 | 999 | 7 | ACA50943 | AcA50943 | Prokaryot | 184 | 14.8 | 77.9 | 444 | 7 | ABQ84702 | AbQ84702 | Manduca s |
| 112 | 15.4 | 81.1 | 1010 | 6 | ABQ28481 | AbQ28481 | Oligonucle | 185 | 14.8 | 77.9 | 447 | 7 | ACA41903 | AcA41903 | Prokaryot |
| 113 | 15.4 | 81.1 | 1010 | 6 | ABQ28480 | Abq28480 | Oligonucle | 186 | 14.8 | 77.9 | 468 | 7 | ABZ38188 | AbZ38188 | N. gonorr |
| 114 | 15.4 | 81.1 | 1035 | 7 | ACA51449 | AcA51449 | Prokaryot | 187 | 14.8 | 77.9 | 468 | 7 | ACA23531 | AcA23531 | Prokaryot |
| 115 | 15.4 | 81.1 | 1035 | 7 | ACA51486 | Aa551486 | Pseudomon | 188 | 14.8 | 77.9 | 468 | 7 | ACA40957 | AcA40957 | Prokaryot |
| 116 | 15.4 | 81.1 | 1176 | 4 | AA551486 | Aa551486 | Pseudomon | 189 | 14.8 | 77.9 | 468 | 7 | ACA26504 | AcA26504 | Prokaryot |
| 117 | 15.4 | 81.1 | 1176 | 7 | ACA51938 | AcA51938 | Prokaryot | 190 | 14.8 | 77.9 | 471 | 7 | ACA41649 | AcA41649 | Prokaryot |
| 118 | 15.4 | 81.1 | 1197 | 4 | AAH51360 | Aah51360 | Mycobacte | 191 | 14.8 | 77.9 | 480 | 5 | AAH88727 | Aah88727 | Cellulose |
| 119 | 15.4 | 81.1 | 1265 | 6 | ABK75033 | AbK75033 | Bacillus | 192 | 14.8 | 77.9 | 480 | 8 | ACL11879 | AcL11879 | DNA clone |
| 120 | 15.4 | 81.1 | 1361 | 6 | ABK75033 | AbK75033 | Bacillus | 193 | 14.8 | 77.9 | 493 | 3 | AAH30710 | Aah30710 | Human col |
| 121 | 15.4 | 81.1 | 1515 | 7 | ACA38048 | AcA38048 | Prokaryot | 194 | 14.8 | 77.9 | 493 | 8 | ACL11939 | AcL11939 | DNA clone |
| 122 | 15.4 | 81.1 | 1576 | 4 | ABL10421 | AbL10421 | Drosophil | 195 | 14.8 | 77.9 | 500 | 6 | ABQ34051 | AbQ34051 | Oligonucle |
| 123 | 15.4 | 81.1 | 1699 | 4 | AA711430 | Aa711430 | Corynebac | 196 | 14.8 | 77.9 | 500 | 6 | ABQ34050 | AbQ34050 | Oligonucle |
| 124 | 15.4 | 81.1 | 1713 | 5 | AAH65969 | Aah65969 | C. Glutami | 197 | 14.8 | 77.9 | 501 | 6 | ABQ28996 | AbQ28996 | Oligonucle |
| 125 | 15.4 | 81.1 | 1836 | 4 | AA711428 | Aa711428 | Corynebac | 198 | 14.8 | 77.9 | 501 | 6 | ABQ28997 | AbQ28997 | Oligonucle |
| 126 | 15.4 | 81.1 | 1923 | 6 | AB214337 | Ab214337 | Arabidops | 199 | 14.8 | 77.9 | 523 | 8 | ACL11933 | AcL11933 | DNA clone |
| 127 | 15.4 | 81.1 | 1928 | 3 | AA795762 | Aa795762 | Arabidops | 200 | 14.8 | 77.9 | 534 | 7 | ABZ39636 | AbZ39636 | N. gonorr |
| 128 | 15.4 | 81.1 | 1928 | 3 | AA652930 | Aa652930 | Arabidops | 201 | 14.8 | 77.9 | 542 | 8 | ACL11946 | AcL11946 | DNA clone |
| 129 | 15.4 | 81.1 | 1942 | 7 | ACA26381 | AcA26381 | Prokaryot | 202 | 14.8 | 77.9 | 551 | 8 | ACL11947 | AcL11947 | DNA clone |
| 130 | 15.4 | 81.1 | 2110 | 3 | AA261243 | Aa261243 | DNA encod | 203 | 14.8 | 77.9 | 553 | 8 | ACL11892 | AcL11892 | DNA clone |
| 131 | 15.4 | 81.1 | 2602 | 4 | ABL26431 | AbL26431 | Drosophil | 204 | 14.8 | 77.9 | 554 | 8 | ACL11905 | AcL11905 | DNA clone |
| 132 | 15.4 | 81.1 | 2832 | 4 | ABL04241 | AbL04241 | Drosophil | 205 | 14.8 | 77.9 | 555 | 7 | ABZ39635 | AbZ39635 | N. gonorr |
| 133 | 15.4 | 81.1 | 2832 | 4 | ABL04243 | AbL04243 | Drosophil | 206 | 14.8 | 77.9 | 558 | 8 | ACL11932 | AcL11932 | DNA clone |
| 134 | 15.4 | 81.1 | 3306 | 2 | AAQ22202 | AaQ22202 | A. chryso | 207 | 14.8 | 77.9 | 558 | 8 | ACL11940 | AcL11940 | DNA clone |
| 135 | 15.4 | 81.1 | 3306 | 2 | AAQ23005 | AaQ23005 | Phosphogl | 208 | 14.8 | 77.9 | 558 | 8 | ACL11897 | AcL11897 | DNA clone |
| 136 | 15.4 | 81.1 | 3483 | 4 | ABL05668 | AbL05668 | Drosophil | 209 | 14.8 | 77.9 | 562 | 8 | ACL11887 | AcL11887 | DNA clone |
| 137 | 15.4 | 81.1 | 3632 | 4 | ABL10420 | AbL10420 | Drosophil | 210 | 14.8 | 77.9 | 562 | 8 | ACL11882 | AcL11882 | DNA clone |
| 138 | 15.4 | 81.1 | 3753 | 4 | ABL05306 | AbL05306 | Drosophil | 211 | 14.8 | 77.9 | 562 | 8 | ACL11882 | AcL11882 | DNA clone |
| 139 | 15.4 | 81.1 | 5019 | 4 | ABL26430 | AbL26430 | Drosophil | 212 | 14.8 | 77.9 | 570 | 7 | ACA40933 | AcA40933 | Prokaryot |
| 140 | 15.4 | 81.1 | 6439 | 9 | ADB78971 | AdB78971 | Tick chit | 213 | 14.8 | 77.9 | 575 | 8 | ACL11885 | AcL11885 | DNA clone |
| 141 | 15.4 | 81.1 | 8372 | 4 | ABL10436 | AbL10436 | Drosophil | 214 | 14.8 | 77.9 | 580 | 8 | ACL11955 | AcL11955 | DNA clone |
| 142 | 15.4 | 81.1 | 8391 | 7 | ACA64735 | AcA64735 | Mycobacte | 215 | 14.8 | 77.9 | 580 | 8 | ACL11934 | AcL11934 | DNA clone |
| 143 | 15.4 | 81.1 | 9210 | 4 | AAH52046 | Aah52046 | Mycobacte | 216 | 14.8 | 77.9 | 582 | 8 | ACL11948 | AcL11948 | DNA clone |
| 144 | 15.4 | 81.1 | 10191 | 4 | ABL04240 | AbL04240 | Drosophil | 217 | 14.8 | 77.9 | 587 | 8 | ACL11870 | AcL11870 | DNA clone |
| 145 | 15.4 | 81.1 | 10283 | 4 | ABL04242 | AbL04242 | Drosophil | 218 | 14.8 | 77.9 | 588 | 8 | ACL11884 | AcL11884 | DNA clone |
| 146 | 15.4 | 81.1 | 110000 | 4 | AAI99682_01 | Continuation (2 of | 219 | 14.8 | 77.9 | 589 | 8 | ACL11937 | AcL11937 | DNA clone | |
| 147 | 15.4 | 81.1 | 110000 | 4 | AAI99682_28 | Continuation (29 o | 220 | 14.8 | 77.9 | 592 | 8 | ACL11918 | AcL11918 | DNA clone | |
| 148 | 15.4 | 81.1 | 110000 | 4 | AAI99683_01 | Continuation (2 of | 221 | 14.8 | 77.9 | 592 | 8 | ACL11888 | AcL11888 | DNA clone | |
| 149 | 15.4 | 81.1 | 110000 | 4 | AAI99683_28 | Continuation (29 o | 222 | 14.8 | 77.9 | 592 | 8 | ACL11864 | AcL11864 | DNA clone | |
| 150 | 15.4 | 81.1 | 110000 | 4 | AAI99683_28 | Continuation (29 o | 223 | 14.8 | 77.9 | 592 | 8 | ACL11893 | AcL11893 | DNA clone | |
| 151 | 15.4 | 81.1 | 34980 | 5 | AAH68527 | AaH68527 | C. Glutami | 224 | 14.8 | 77.9 | 603 | 8 | ACL11895 | AcL11895 | DNA clone |
| 152 | 15.4 | 81.1 | 34980 | 6 | ABQ81847 | AbQ81847 | Bifidobac | 225 | 14.8 | 77.9 | 603 | 8 | ACL11876 | AcL11876 | DNA clone |
| 153 | 15.4 | 81.1 | 34980 | 6 | ABQ81848 | AbQ81848 | Bifidobac | 226 | 14.8 | 77.9 | 607 | 8 | ACL11913 | AcL11913 | DNA clone |
| 154 | 15.4 | 81.1 | 34980 | 6 | ABQ81844 | AbQ81844 | Bifidobac | 227 | 14.8 | 77.9 | 611 | 8 | ACL11928 | AcL11928 | DNA clone |
| 155 | 15.4 | 78.9 | 348 | 6 | ABN24546 | Abn24546 | Human ORF | 228 | 14.8 | 77.9 | 612 | 3 | AAZ54455 | Aaz54455 | Neisseria |
| 156 | 15.4 | 78.9 | 482 | 8 | ACH29222 | ACH29222 | Human adu | 229 | 14.8 | 77.9 | 612 | 8 | ACL11924 | AcL11924 | DNA clone |
| 157 | 15.4 | 78.9 | 1001 | 9 | ADC03458 | AdC03458 | Wheat flo | 230 | 14.8 | 77.9 | 612 | 8 | ACL11930 | AcL11930 | DNA clone |
| 158 | 15.4 | 78.9 | 1189 | 3 | AAZ58264 | Aaz58264 | Corn cell | 231 | 14.8 | 77.9 | 614 | 8 | ACL11957 | AcL11957 | DNA clone |
| 159 | 15.4 | 78.9 | 1632 | 7 | ABZ37939 | AbZ37939 | N. gonorr | 232 | 14.8 | 77.9 | 615 | 8 | ACL11890 | AcL11890 | DNA clone |
| 160 | 15.4 | 78.9 | 1817 | 7 | ACA41325 | AcA41325 | Prokaryot | 233 | 14.8 | 77.9 | 616 | 8 | ACL11894 | AcL11894 | DNA clone |
| 161 | 15.4 | 78.9 | 1947 | 4 | ABL05555 | AbL05555 | Prokaryot | 234 | 14.8 | 77.9 | 618 | 8 | ACL11949 | AcL11949 | DNA clone |
| 162 | 15.4 | 78.9 | 1977 | 2 | AAZ31126 | Aaz31126 | Corn suc | 235 | 14.8 | 77.9 | 619 | 8 | ACL11943 | AcL11943 | DNA clone |
| 163 | 15.4 | 78.9 | 1983 | 7 | ACA50935 | AcA50935 | Prokaryot | 236 | 14.8 | 77.9 | 620 | 8 | ACL11883 | AcL11883 | DNA clone |
| 164 | 15.4 | 78.9 | 1983 | 7 | ACA51438 | AcA51438 | Prokaryot | 237 | 14.8 | 77.9 | 621 | 8 | ACL11898 | AcL11898 | DNA clone |
| 165 | 15.4 | 78.9 | 2961 | 8 | ADA14469 | Ada14469 | Mouse spe | 238 | 14.8 | 77.9 | 622 | 8 | ACL11915 | AcL11915 | DNA clone |
| 166 | 15.4 | 78.9 | 4159 | 4 | ABL05554 | AbL05554 | Drosophil | 239 | 14.8 | 77.9 | 624 | 8 | ACL11942 | AcL11942 | DNA clone |
| 167 | 15.4 | 78.9 | 10198 | 5 | AA61282 | Aa61282 | N. magada | 240 | 14.8 | 77.9 | 626 | 8 | ACL11952 | AcL11952 | DNA clone |
| 168 | 15.4 | 78.9 | 77536 | 3 | AAAL4651 | Aaal4651 | Nucleotid | 241 | 14.8 | 77.9 | 628 | 8 | ACL11935 | AcL11935 | DNA clone |
| 169 | 15.4 | 78.9 | 349980 | 6 | ABQ81849 | AbQ81849 | Bifidobac | 242 | 14.8 | 77.9 | 630 | 8 | ACL11902 | AcL11902 | DNA clone |

c 243 14.8 77.9 632 8 ACL11867
c 244 14.8 77.9 632 8 ACL11931
c 245 14.8 77.9 632 8 ACL11914
c 246 14.8 77.9 632 8 ACL11941
c 247 14.8 77.9 635 8 ACL11925
c 248 14.8 77.9 636 8 ACL11866
c 249 14.8 77.9 639 8 ACL11877
c 250 14.8 77.9 642 8 ACL11951
c 251 14.8 77.9 647 8 ACL11926
c 252 14.8 77.9 651 8 ACL11873
c 253 14.8 77.9 651 8 ACL11878
c 254 14.8 77.9 656 8 ACL11921
c 255 14.8 77.9 657 8 ACL11927
c 256 14.8 77.9 664 8 ACL11919
c 257 14.8 77.9 665 8 ACL11938
c 258 14.8 77.9 668 7 ABZ53865
c 259 14.8 77.9 674 8 ACL11936
c 260 14.8 77.9 678 8 ACL11916
c 261 14.8 77.9 681 8 ACL11865
c 262 14.8 77.9 688 8 ACL11896
c 263 14.8 77.9 694 8 ACL11891
c 264 14.8 77.9 698 8 ACL11871
c 265 14.8 77.9 700 4 RAH92000
c 266 14.8 77.9 720 8 ACL11868
c 267 14.8 77.9 726 8 ADA48757
c 268 14.8 77.9 735 7 ACA43973
c 269 14.8 77.9 741 4 AAS54206
c 270 14.8 77.9 741 7 ACA42468
c 271 14.8 77.9 741 7 ACA40915
c 272 14.8 77.9 787 3 AAS58916
c 273 14.8 77.9 789 5 AAH65192
c 274 14.8 77.9 789 7 ACA01924
c 275 14.8 77.9 818 8 ADA48139
c 276 14.8 77.9 819 7 ACF39432
c 277 14.8 77.9 819 7 ACA38666
c 278 14.8 77.9 822 7 ACA40868
c 279 14.8 77.9 825 2 AAQ11855
c 280 14.8 77.9 850 3 AAF14891
c 281 14.8 77.9 864 7 ADA69341
c 282 14.8 77.9 874 6 ABQ18184
c 283 14.8 77.9 874 6 ABQ18185
c 284 14.8 77.9 879 7 ACA23319
c 285 14.8 77.9 885 6 ABK73038
c 286 14.8 77.9 902 6 ABQ21053
c 287 14.8 77.9 902 6 ABQ21052
c 288 14.8 77.9 918 6 ABQ90033
c 289 14.8 77.9 918 6 ABQ90034
c 290 14.8 77.9 933 4 AAH43317
c 291 14.8 77.9 933 9 AAD59980
c 292 14.8 77.9 954 6 ABA91410
c 293 14.8 77.9 966 6 ABN89590
c 294 14.8 77.9 966 7 ACA42108
c 295 14.8 77.9 966 7 ACA37709
c 296 14.8 77.9 987 5 AAH68462
c 297 14.8 77.9 993 9 ADC23846
c 298 14.8 77.9 1005 3 AAF14888
c 299 14.8 77.9 1009 5 AAH8784
c 300 14.8 77.9 1012 6 ABK73256

ALIGNMENTS

RESULT 1
AAT49174
ID AAT49174 standard; DNA; 393 BP.

XX
AC AAT49174;
XX

DT 16-OCT-1997 (first entry)

XX
XX Partial DNA clone HinP1-27 encoding immunostimulatory peptide.
DE
XX

KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX WO9700067-A1.
XX
PD 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US010375.
XX
XX 15-JUN-1995; 95US-0000254P.
XX
XX (UYVI-) UNIV VICTORIA.
XX Nano FE;
XX WPI; 1997-077347/07.
XX
XX New immuno-stimulatory peptide(s) of Mycobacterium tuberculosis - useful
in vaccines, diagnostic skin test, immunoassay and gene isolation.
XX
XX Claim 1; Page 1; 79pp; English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
which encode partial sequences of immunostimulatory peptides. Each of the
clones encode at least one immunostimulatory T cell epitope. The clones
were identified by testing over 300 fusion clones (alkaline phosphatase-
M. tuberculosis peptide fusions) for their ability to stimulate
interferon (IFN)-gamma production. 80 clones were initially designated to
have some ability to stimulate IFN-gamma production, of which 76 are
shown in AAT49100-175. These sequences can be used to obtain the full
length M. tuberculosis genes and corresponding proteins using standard
techniques. The peptides are useful in vaccines, as reagents in an
improved tuberculin skin test (especially using peptides different from
those used in vaccines so as to allow differentiation between vaccinated
and infected subjects), and as immunoassay reagents for detecting specific
antibodies. An advantage of these peptides is that they stimulate
production of IFN-gamma (critical for a protective immune response to M.
tuberculosis) by CD4-positive T cells
XX
SQ Sequence 393 BP; 72 A; 120 C; 119 G; 82 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCATGACC 19
Db 12 GTCGACGCCCGCATGACC 30
|||||
|||

RESULT 2
AAT99682_42/c
Continuation (43 of 45) of AAT99682 from base 4200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAT99682 Accession AAT99682
WP Fragment Name Begin End
WP AAT99682_00 1 110000
WP AAT99682_01 100001 210000
WP AAT99682_02 200001 310000
WP AAT99682_03 300001 410000
WP AAT99682_04 400001 510000
WP AAT99682_05 500001 610000
WP AAT99682_06 600001 710000
WP AAT99682_07 700001 810000
WP AAT99682_08 800001 910000
WP AAT99682_09 900001 1010000
WP AAT99682_10 1000001 1110000
WP AAT99682_11 1100001 1210000
WP AAT99682_12 1200001 1310000
WP AAT99682_13 1300001 1410000
WP AAT99682_14 1400001 1510000

| | | | |
|----|------------|---------|---------|
| WP | AAI9682_15 | 1500001 | 1610000 |
| WP | AAI9682_16 | 1600001 | 1710000 |
| WP | AAI9682_17 | 1700001 | 1810000 |
| WP | AAI9682_18 | 1800001 | 1910000 |
| WP | AAI9682_19 | 1900001 | 2010000 |
| WP | AAI9682_20 | 2000001 | 2110000 |
| WP | AAI9682_21 | 2100001 | 2210000 |
| WP | AAI9682_22 | 2200001 | 2310000 |
| WP | AAI9682_23 | 2300001 | 2410000 |
| WP | AAI9682_24 | 2400001 | 2510000 |
| WP | AAI9682_25 | 2500001 | 2610000 |
| WP | AAI9682_26 | 2600001 | 2710000 |
| WP | AAI9682_27 | 2700001 | 2810000 |
| WP | AAI9682_28 | 2800001 | 2910000 |
| WP | AAI9682_29 | 2900001 | 3010000 |
| WP | AAI9682_30 | 3000001 | 3110000 |
| WP | AAI9682_31 | 3100001 | 3210000 |
| WP | AAI9682_32 | 3200001 | 3310000 |
| WP | AAI9682_33 | 3300001 | 3410000 |
| WP | AAI9682_34 | 3400001 | 3510000 |
| WP | AAI9682_35 | 3500001 | 3610000 |
| WP | AAI9682_36 | 3600001 | 3710000 |
| WP | AAI9682_37 | 3700001 | 3810000 |
| WP | AAI9682_38 | 3800001 | 3910000 |
| WP | AAI9682_39 | 3900001 | 4010000 |
| WP | AAI9682_40 | 4000001 | 4110000 |
| WP | AAI9682_41 | 4100001 | 4210000 |
| WP | AAI9682_42 | 4200001 | 4310000 |
| WP | AAI9682_43 | 4300001 | 4410000 |
| WP | AAI9682_44 | 4400001 | 4411529 |

Query Match
Best Local Similarity 100.0%;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Score 19; DB 4; Length 110000;
Pred. No. 38;

QY 1 GTGACGCCGGCGATGACC 19
|||||

DB 27060 GTGACGCCGGCGATGACC 27042

RESULT 3
AAI9683_42/c
Continuation (43 of 44) of AAI9683 from base 4200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI9683 Accession Aai9683

Fragment Name Begin End

WP AAI9683_00 1 110000

WP AAI9683_01 100001 210000

WP AAI9683_02 200001 310000

WP AAI9683_03 300001 410000

WP AAI9683_04 400001 510000

WP AAI9683_05 500001 610000

WP AAI9683_06 600001 710000

WP AAI9683_07 700001 810000

WP AAI9683_08 800001 910000

WP AAI9683_09 900001 1010000

WP AAI9683_10 1000001 1110000

WP AAI9683_11 1100001 1210000

WP AAI9683_12 1200001 1310000

WP AAI9683_13 1300001 1410000

WP AAI9683_14 1400001 1510000

WP AAI9683_15 1500001 1610000

WP AAI9683_16 1600001 1710000

WP AAI9683_17 1700001 1810000

WP AAI9683_18 1800001 1910000

WP AAI9683_19 1900001 2010000

WP AAI9683_20 2000001 2110000

WP AAI9683_21 2100001 2210000

WP AAI9683_22 2200001 2310000

WP AAI9683_23 2300001 2410000

WP AAI9683_24 2400001 2510000

WP AAI9683_25 2500001 2610000

WP AAI9683_26 2600001 2710000

WP AAI9683_27 2700001 2810000

Query Match
Best Local Similarity 100.0%;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Score 19; DB 4; Length 110000;
Pred. No. 38;

QY 1 GTGACGCCGGCGATGACC 19
|||||

DB 27060 GTGACGCCGGCGATGACC 27042

RESULT 4
AAI37969
ID ACI37969 standard; DNA; 25 BP.

XX ACI37969;

DT 13-OCT-2003 (first entry)

DE Human microarray DNA oligonucleotide SEQ ID NO 37960.

KW EST; ss; probe; expressed sequence tag; microarray; gene expression;
Genetic variation; allelic marker; polymorphism; human;
cross-species comparison.

OS Homo sapiens.

XX US2003104410-A1.
05-JUN-2003.
15-MAR-2002; 2002US-00098263.
16-MAR-2001; 2001US-0276759P.
(AFFY-) AFFYMETRIX INC.
Mittmann MP;
WPI; 2003-567953/53.

New array of nucleic acid probes, useful for in situ hybridization, in Southern, Northern or dot-blot hybridization to identify or detect the sequence or specific mutations of any gene.

Claim 1; SEQ ID NO 37960; 9pp; English.

The invention discloses a microarray comprising a plurality of nucleic acid probes including one of 2,018,500 fully defined sequences, or its perfect match, antisense match or antisense mismatch. The array is used in monitoring gene expression levels by hybridisation to a DNA library, in analysis of genetic variation or in hybridisation of tag-labelled compounds. The nucleic acid probes are specifically designed for analysis of at least one target sequence. The method of analysis comprises hybridising at least one or more nucleic acids to at least two or more nucleic acid probes and detecting the hybridisation. The nucleic acid probes are attached to a solid support. The analysis comprises monitoring

CC gene expression levels, identifying biallelic markers or polymorphisms,
 CC or family members of a gene and a cross-species comparison. Each of the
 CC nucleic acids further comprises a tag sequence. The array of nucleic acid
 CC probes is useful in situ hybridisation, in Southern, Northern or dot-
 CC blot hybridisation to identify or detect the sequence or specific
 CC mutations of any gene, in mapping the 5' termini of mRNA molecules by
 CC primer extensions or in screening cDNA or genomic libraries or subclones
 CC for additional subclones containing segments of DNA that have been
 CC isolated and previously sequenced. The sequence presented is one of the
 CC nucleic acid probes incorporated in the microarray. Note: The sequence
 CC data for this patent can also be obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html
 XX
 SQ Sequence 25 BP; 4 A; 8 C; 9 G; 4 T; 0 U; 0 Other;

Query Match 89.5%; Score 17; DB 8; Length 25;
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCACGCCGCCGATGA 17
 |||||
 Db 9 GTCACGCCGCCGATGA 25

RESULT 5

ABL72618/c
 ID ABL72618 standard; cDNA; 270 BP.

XX ABL72618;

XX 14-MAY-2002 (first entry)

DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:1992.

XX Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPS;
 KW inheritance; characteristic; growth; development; disease resistance;
 KW environmental adaptability; quality; yield; molecular marker;
 KW multigene trait; plant breeding; corn tassel; gene; ss.

XX Zea mays.

XX US2001051335-A1.

XX 13-DEC-2001.

XX 16-APR-1999; 99US-00294093.

XX 21-APR-1998; 99US-0082567P.

XX (LALG/) LALGUDI R V.

XX (ITOL/) ITO L Y.

XX (SHER/) SHERMAN B K.

XX Lalgudi RV, Ito LY, Sherman BK;

XX WPI; 2002-163647/21.

DR Novel purified corn tassel-derived polynucleotide useful for determining
 PT altered gene expression, to recover regulatory elements and to follow
 PT inheritance of desirable characteristics through hybrid breeding
 PT programs.

XX Claim 1; SEQ ID NO 1992; 201pp; English.

XX The present sequence describes a purified corn tassel-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (1)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (1) are also useful in
 CC the evaluation, and alteration of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability,

CC quality and yield, and as molecular markers for studying inheritance of
 CC multigene traits in a plant breeding program. (1) can be used to produce
 CC a tassel-specific profile of gene transcription, a transcript image, to
 CC clone regulatory elements for use in transformation vectors, to express a
 CC polypeptide, to identify, isolate or extend identical or related corn
 CC tassel nucleic acid sequences from DNA libraries, in nucleic acid
 CC hybridisation or amplification technologies, as query sequences to
 CC determine homology of known sequences, as probe for use in Southern or
 CC Northern hybridisation, and to identify the presence of and/or to
 CC determine the degree of similarity between two (or more) nucleic acid
 CC sequences

XX
 SQ Sequence 270 BP; 40 A; 97 C; 82 G; 50 T; 0 U; 1 Other;

Query Match 86.3%; Score 16.4; DB 6; Length 270;

Best Local Similarity 94.4%; Pred. No. 7.2e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
 |||||
 Db 87 GTCGACGCCGCCGATGAC 70

RESULT 6

AAC39698/c

ID AAC39698 standard; DNA; 432 BP.

XX AAC39698;

XX 17-OCT-2000 (first entry)

DE Zea mays DNA fragment SEQ ID NO: 25575.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway; metabolic; pathway;

XX promoter; termination sequence; corn; ss.

XX Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 03-MAR-1999; 99US-0123180P.

XX 23-MAR-1999; 99US-0123548P.

XX 25-MAR-1999; 99US-0125788P.

XX 01-APR-1999; 99US-0126264P.

XX 06-APR-1999; 99US-0127462P.

XX 08-APR-1999; 99US-0128234P.

XX 16-APR-1999; 99US-0128714P.

XX 19-APR-1999; 99US-0130077P.

XX 21-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 28-APR-1999; 99US-0130891P.

XX 30-APR-1999; 99US-0131449P.

XX 30-APR-1999; 99US-0132048P.

XX 04-MAY-1999; 99US-0132407P.

XX 05-MAY-1999; 99US-0132484P.

XX 06-MAY-1999; 99US-0132485P.

XX 07-MAY-1999; 99US-0132487P.

XX 11-MAY-1999; 99US-0132863P.

XX 14-MAY-1999; 99US-0134218P.

XX 14-MAY-1999; 99US-0134219P.

XX 14-MAY-1999; 99US-0134221P.

XX 14-MAY-1999; 99US-0134370P.

XX 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137503P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 14-JUN-1999; 99US-0138847P.
 PR 16-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 21-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0143977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 26-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149375P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 23-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154019P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 05-OCT-1999; 99US-0157117P.
 PR 08-OCT-1999; 99US-0157753P.
 PR 08-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 23-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.

```
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      86.3%; Score 16.4; DB 3; Length 432;
Best Local Similarity 94.4%; Pred. NO. 7+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
      |||||
Db      421 TCGACGCCGCCGATGACC 404

RESULT 7
AAC41047/c
ID AAC41047 standard; DNA; 445 BP.
XX
AC AAC41047;
XX
DT 17-OCT-2000 (first entry)
DE XX
DE Zea mays DNA fragment SEQ ID NO: 30455.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic; pathway;
KW promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 29-MAR-1999; 99US-0125268P.
PR 29-MAR-1999; 99US-0125785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 15-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142330P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
```

```

PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 23-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159337P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161559P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      86.3%; Score 16.4; DB 3; Length 445;
Best Local Similarity 94.4%; Pred. No. 7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19
Db 436 TCGACGCCGCCGATGACC 419

RESULT 8
ADA70332/C
ID ADA70332 standard; DNA; 963 BP.
XX AC ADA70332;
XX DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 3655.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX KW gene; ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX FA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX FI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX DR WPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
XX PT pathogenic infection for conferring resistance or tolerance to a plant to
XX PT bacterial, fungal or viral infection by determining or detecting plant
XX PT gene expression.
XX PS Claim 6; SEQ ID NO 3655; 899pp; English.
XX CC The present invention relates to a method (M1) for identifying genes
XX CC involved in plant resistance or response to pathogenic infection. M1
XX CC comprises identifying a gene whose expression is significantly altered in
XX CC the incompatible interaction of plant gene expression relative to
XX CC expression of the gene in an uninfected plant, in a mutant plant that
XX CC does not express a gene associated with response to pathogenic infection,
XX CC or in a corresponding incompatible or compatible interaction. (M1) is
XX CC useful for conferring resistance to resistance or tolerance to a plant to
XX CC bacterial, fungal or viral infection. The present sequence was used to
XX CC illustrate the invention.
SQ Sequence 963 BP; 159 A; 334 C; 292 G; 178 T; 0 U; 0 Other;

Query Match      86.3%; Score 16.4; DB 7; Length 963;
Best Local Similarity 94.4%; Pred. No. 6.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 501 GTCGACGCCGCCGATGAC 484

RESULT 9
ADA48367/C
ID ADA48367 standard; DNA; 1077 BP.
XX AC ADA48367;

```

```

XX 20-NOV-2003 (first entry)
DT Rice gene conferring disease resistance in plants.
XX disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
XX Oryza sativa.
XX WO2003000906-A2.
XX 03-JAN-2003.
XX 21-JUN-2002; 2002WO-IB002453.
XX 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0322277P.
PR 22-MAR-2002; 2002US-0366535P.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Katagiri F, Krepis J, Provart N, Ricke D, Zhu T;
XX WPI; 2003-184052/18.
DR P-PSDB; ADA48368.
XX New polynucleotide comprising a plant nucleotide sequence having an open
PT reading frame that encodes a polypeptide associated with disease
PT resistance, useful for conferring resistance or tolerance to a plant
PT pathogen.
XX Claim 1; SEQ ID NO 437; 299pp; English.
XX The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a gene
CC conferring disease resistance used in the invention.
XX Sequence 1077 BP; 189 A; 368 C; 363 G; 157 T; 0 U; 0 Other;
SQ Query Match 86.3%; Score 16.4; DB 8; Length 1077;
Best Local Similarity 94.4%; Pred. NO. 6.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTCGACGCCGCCGATGAC 18
Db 171 GTCGACGCCGCCGATGAC 154

RESULT 10
AAD29838/c
ID AAD29838 standard; DNA; 1560 BP.
XX AAD29838;
XX 17-MAY-2002 (first entry)
XX Rice NADPH dependent thioredoxin reductase DNA (NTR).
XX Transgenic plant; thioredoxin reductase; starch; protein; grain;
KW milling process; enzyme; rice; ds.
XX Oryza sativa.
XX Key Location/Qualifiers
PH 406..1338
FT CDS /*tag=a
FT /*product="Rice NTR"
FT FT
XX

```

```

PN WO200198509-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-EP006918.
XX 21-JUN-2000; 2000US-00598747.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Lanahan MB, Desai NM, Gasdaska PY;
PI WPI; 2002-179557/23.
XX P-PSDB; AAE18733.
XX Transgenic plant coding for eukaryotic thioredoxin reductase at elevated
PT levels useful for separating the starch and protein components of grain
PT in a milling process.
XX Claim 12; Page 81; 86pp; English.
XX The present invention relates to a transgenic plant comprising
CC heterologous DNA coding for eukaryotic thioredoxin reductase integrated
CC into its nuclear or plastid genome and use of thioredoxin reductase for
CC separating the starch and protein components of grain in a milling
CC process. Transgenic plant is used for separating the starch and protein
CC components of grain in a milling process. Transgenic plant may be used to
CC produce thioredoxin reductase at elevated levels. Delivery of thioredoxin
CC reductase eliminates the need to develop exogenous sources for addition
CC during processing. Secondly, physical disruption of seed integrity is not
CC necessary to bring the enzyme in contact with the storage or matrix
CC proteins of the seed prior to processing or as an extra processing step.
CC The present sequence is rice NADPH dependent thioredoxin reductase DNA
CC (NTR)
XX Sequence 1560 BP; 302 A; 435 C; 439 G; 372 T; 0 U; 12 Other;
SQ Query Match 86.3%; Score 16.4; DB 6; Length 1560;
Best Local Similarity 94.4%; Pred. NO. 6.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTCGACGCCGCCGATGAC 18
Db 906 GTCGACGCCGCCGATGAC 889

RESULT 11
ACA27231/c
ID ACA27231 standard; DNA; 1948 BP.
XX ACA27231;
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #8888.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Bordetella pertussis.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX

```

PA (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR P-PSDB; ABU23361.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 14; SEQ ID NO 15101; 1766pp; English.
 PS
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1948 BP; 289 A; 684 C; 682 G; 293 T; 0 U; 0 Other;
 Query Match 86.3%; Score 16.4; DB 7; Length 1948;
 Best Local Similarity 94.4%; Pred. No. 6.3e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCGACGCCCGCGATGAC 18
 Db 96 GTCCGCGCGCGCGATGAC 79
 RESULT 12
 ABZ16341
 ID ABZ16341 standard; DNA; 2000 BP.
 XX
 XX ABZ16341;
 XX
 XX 21-JAN-2003 (first entry)
 DT
 XX Arabidopsis thaliana stress regulated gene SEQ ID NO 4146.
 DE
 XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 KW
 XX Arabidopsis thaliana.
 OS
 XX Arabidopsis thaliana.
 XX
 XX WO200216655-A2.
 PN
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;

PD 28-FEB-2002.
 XX
 XX 24-AUG-2001; 2001WO-US026685.
 PF
 XX 24-AUG-2000; 2000US-0227866P.
 PR
 XX 26-JAN-2001; 2001US-0264647P.
 PR
 XX 22-JUN-2001; 2001US-0300111P.
 XX
 XX (SCRI) SCRIPPS RES INST.
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 PA
 PI Harper JF, Kreps J, Wang X, Zhu T;
 PI
 XX WPI; 2002-304127/34.
 DR
 XX Identifying a stress condition to which a plant cell has been exposed and
 PT producing plants with increased tolerance to these abiotic stresses.
 PT
 XX Claim 14; SEQ ID NO 4146; 577pp + Sequence Listing; English.
 PS
 XX The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising: (a) contacting nucleic acid
 CC representative of expressed polynucleotides in the plant cell with an
 CC array or probes representative of the plant cell genome; and (b)
 CC detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification but is based on sequence
 CC information supplied to Derwent by the European Patent Office
 XX
 SQ Sequence 2000 BP; 683 A; 368 C; 282 G; 667 T; 0 U; 0 Other;
 Query Match 86.3%; Score 16.4; DB 6; Length 2000;
 Best Local Similarity 94.4%; Pred. No. 6.3e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCGACGCCCGCGATGAC 18
 Db 353 GTCCGCGCGCGCGATGAC 370
 RESULT 13
 AAA81465
 ID AAA81465 standard; DNA; 47475 BP.
 XX
 XX AAA81465;
 AC
 XX 04-DEC-2000 (first entry)
 DT
 XX N. meningitidis partial DNA sequence gnm_13 SEQ ID NO:13.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX WO200022430-A2.
 FN
 XX 20-APR-2000.
 PD
 XX 08-OCT-1999; 99WO-US023573.
 PF
 XX 09-OCT-1998; 98US-0103794P.
 PR
 XX 30-APR-1999; 99US-0132068P.
 PR
 XX (CHIR) CHIRON CORP.
 XX
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;

PI Rappuoli R, Pizza M;
XX WPI; 2000-318079/27.
DR
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used
PT in the diagnosis and treatment of *N. meningitidis* infection and other
PT *Neisseria* infections, for example, *N. gonorrhoea*.
PS
XX Claim 7; Page 383-397; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic proteins
CC from *Neisseria* genomic sequences. AA81453 to AA82414 represent
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;
CC AA81260 to AA81303 and AA825620 to AA825663 represent *Neisseria* DNA
CC sequences and their corresponding proteins; AA81254 to AA81259 and
CC AA81304 to AA81321 represent PCR primers used in the isolation of
CC *Neisseria meningitidis* DNA sequences; and AA81322 to AA81452 represent
CC *Neisseria meningitidis* Mena polynucleotide ORF sequences, which are all
CC used in the exemplification of the present invention. The nucleic acid
CC sequences, protein sequences, and antibodies against them, can be used in
CC the manufacture of a composition. The composition can be used as a
CC medicament (or in the manufacture of a medicament) for treating,
CC preventing or diagnosing infection due to *Neisseria* bacteria. For
CC example, some of the identified proteins could be components of vaccines
CC against *Meningococcus B*; against all serotypes; and/or against all
CC pathogenic *Neisseriae*. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious *Meningococcus B*
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions
XX
SQ Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 0 U; 1 Other;
Query Match 86.3%; Score 16.4; DB 3; Length 47475;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCGACGCCCGCGATGACC 19
DB 13955 TCGACGCCCGCGAGACC 13972
RESULT 14
AA81489_5
Continuation (6 of 9) of AA81489 from base 500001 (*N. meningitidis* partial DNA sequence
WP Sequence split into 9 fragments LOCUS AA81489 Accession AA81489
WP Fragment Name Begin End
WP AA81489_0 1 110000
WP AA81489_1 100001 210000
WP AA81489_2 200001 310000
WP AA81489_3 300001 410000
WP AA81489_4 400001 510000
WP AA81489_5 500001 610000
WP AA81489_6 600001 710000
WP AA81489_7 700001 810000
WP AA81489_8 800001 837096
Query Match 86.3%; Score 16.4; DB 3; Length 110000;
Best Local Similarity 94.4%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCGACGCCCGCGATGACC 19
DB 23196 TCGACGCCCGGAGACC 23213
RESULT 15
AA199682_02

Continuation (3 of 45) of AA199682 from base 200001 (*Mycobacterium tuberculosis* strain H
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682
WP Fragment Name Begin End
WP AA199682_00 1 110000
WP AA199682_01 100001 210000
WP AA199682_02 200001 310000
WP AA199682_03 300001 410000
WP AA199682_04 400001 510000
WP AA199682_05 500001 610000
WP AA199682_06 600001 710000
WP AA199682_07 700001 810000
WP AA199682_08 800001 910000
WP AA199682_09 900001 1010000
WP AA199682_10 1000001 1110000
WP AA199682_11 1100001 1210000
WP AA199682_12 1200001 1310000
WP AA199682_13 1300001 1410000
WP AA199682_14 1400001 1510000
WP AA199682_15 1500001 1610000
WP AA199682_16 1600001 1710000
WP AA199682_17 1700001 1810000
WP AA199682_18 1800001 1910000
WP AA199682_19 1900001 2010000
WP AA199682_20 2000001 2110000
WP AA199682_21 2100001 2210000
WP AA199682_22 2200001 2310000
WP AA199682_23 2300001 2400001
WP AA199682_24 2400001 2510000
WP AA199682_25 2500001 2610000
WP AA199682_26 2600001 2710000
WP AA199682_27 2700001 2810000
WP AA199682_28 2800001 2910000
WP AA199682_29 2900001 3010000
WP AA199682_30 3000001 3110000
WP AA199682_31 3100001 3210000
WP AA199682_32 3200001 3310000
WP AA199682_33 3300001 3410000
WP AA199682_34 3400001 3510000
WP AA199682_35 3500001 3610000
WP AA199682_36 3600001 3710000
WP AA199682_37 3700001 3810000
WP AA199682_38 3800001 3910000
WP AA199682_39 3900001 4010000
WP AA199682_40 4000001 4110000
WP AA199682_41 4100001 4210000
WP AA199682_42 4200001 4310000
WP AA199682_43 4300001 4410000
WP AA199682_44 4400001 4411529
Query Match 86.3%; Score 16.4; DB 4; Length 110000;
Best Local Similarity 94.4%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCGACGCCCGCGATGAC 18
DB 77268 GACGACGCCCGCGATGAC 77285
RESULT 16
AA199683_02
Continuation (3 of 44) of AA199683 from base 200001 (*Mycobacterium tuberculosis* strain H
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683
WP Fragment Name Begin End
WP AA199683_00 1 110000
WP AA199683_01 100001 210000
WP AA199683_02 200001 310000
WP AA199683_03 300001 410000
WP AA199683_04 400001 510000
WP AA199683_05 500001 610000
WP AA199683_06 600001 710000
WP AA199683_07 700001 810000
WP AA199683_08 800001 910000
WP AA199683_09 900001 1010000

WP AAI99683_10 1110000
 WP AAI99683_11 1210000
 WP AAI99683_12 1310000
 WP AAI99683_13 1410000
 WP AAI99683_14 1510000
 WP AAI99683_15 1610000
 WP AAI99683_16 1710000
 WP AAI99683_17 1810000
 WP AAI99683_18 1910000
 WP AAI99683_19 2010000
 WP AAI99683_20 2110000
 WP AAI99683_21 2210000
 WP AAI99683_22 2310000
 WP AAI99683_23 2410000
 WP AAI99683_24 2510000
 WP AAI99683_25 2610000
 WP AAI99683_26 2710000
 WP AAI99683_27 2810000
 WP AAI99683_28 2910000
 WP AAI99683_29 3010000
 WP AAI99683_30 3110000
 WP AAI99683_31 3210000
 WP AAI99683_32 3310000
 WP AAI99683_33 3410000
 WP AAI99683_34 3510000
 WP AAI99683_35 3610000
 WP AAI99683_36 3710000
 WP AAI99683_37 3810000
 WP AAI99683_38 3910000
 WP AAI99683_39 4010000
 WP AAI99683_40 4110000
 WP AAI99683_41 4210000
 WP AAI99683_42 4310000
 WP AAI99683_43 4403765

Query Match 86.3%; Score 16.4; DB 4; Length 110000;

Best Local Similarity 94.4%; Pred. No. 4.9e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGAC 18
 Db 77380 GACGACGCCGCGATGAC 77397

RESULT 17
 ID AAF21612
 AC AAF21612;
 DT 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.

KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.

XX Neisseria meningitidis.

OS WO200066791-A1.

FN 09-NOV-2000.

PD 08-MAR-2000; 2000WO-US005928.

PF 30-APR-1999; 99US-0132068P.

PR 08-OCT-1999; 99WO-US023573.

PR 28-FEB-2000; 2000GB-00004695.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Frazer CM, Grandi G;
 XX WPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading
 frames are used to detect, treat and prevent Neisserial infections.

Claim 7; Appendix A; 692pp; English.

The present invention describes the full length genome of Neisseria
 meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
 represent fragments of the NMB genomic sequence, as the sequence was too
 long to go in a record on its own it was split into 8 sequences which
 overlap each other at the beginning and end of each sequence by 49980 bp
 (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
 AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
 AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
 given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
 primers which are used in the exemplification of the present invention.
 The NMB genome and fragments from it have antibacterial activity, and can
 be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
 and/or antibodies which binds to the proteins can be used in compositions
 for treating or preventing infection due to Neisserial bacteria or as a
 diagnostic reagent for detecting the presence of Neisserial bacteria or
 of antibodies raised to Neisserial bacteria. Computers, computer memory,
 computer storage medium or computer databases can be used in a search to
 identify open reading frames (ORFs) or coding sequences within the NMB
 genome. The DNA sequences provide further opportunities to find antigenic
 or immunogenic proteins which are more effective in vaccines than the
 outer membrane proteins currently used

Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 3; Length 349980;

Best Local Similarity 94.4%; Pred. No. 4.5e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCGATGACC 19
 Db 158425 TCGACGCCGCGAGACC 158442

RESULT 18

ABQ81846

ID ABQ81846 standard; DNA; 349980 BP.

AC ABQ81846;

DT 19-NOV-2002 (first entry)

XX Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1102.

KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KW antiarrhythmic; antibacterial; inhibitor of salmonella; detection;
 KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX Bifidobacterium longum.

OS Synthetic.

FN EP1227152-A1.

PD 31-JUL-2002.

PF 30-JAN-2001; 2001EP-00102050.

PR 30-JAN-2001; 2001EP-00102050.

XX (NEST) SOC PROD NESTLE SA.

PA WPI; 2002-668397/72.

XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT a probe or primer for detecting and/or identifying Bifidobacterium longum
PT in a biological sample.
XX
XX Disclosure; SEQ ID NO 1102; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in AB081842 and AB081843, or a sequence exhibiting at
CC least 90% identity or which hybridizes with the sequences given in
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding a
CC fusion protein, comprising a sequence selected from 1097 sequences given
CC in ABP65258 to ABP65354 ligated in frame to a polynucleotide encoding a
CC heterologous polypeptide. (I) has anti-diarrheic and antibacterial
CC activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC a probe) is useful for the detection and/or identification of
CC Bifidobacterium longum in a biological sample. A carrier containing the
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be
CC used for preventing and/or treating diarrhoea brought about by pathogenic
CC bacteria and/or rotavirus. The carrier is a food composition selected
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC products, ice-creams, fermented cereal based products, milk based
CC powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC Listing from the present invention but not mentioned further within the
CC specification. N.B. The sequence data for this patent is not represented
CC in the printed specification but is based on sequence information
CC supplied by the European Patent Office
XX
XX SQ Sequence 349980 BP; 69200 A; 103414 C; 105147 G; 72219 T; 0 U; 0 Other;
Query Match 86.3%; Score 16.4; DB 6; Length 349980;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCGACGCCGCCGATG 18
DB 269605 GTCGACGCCGCCGATG 269622
RESULT 19
ADC23768/c
ID ADC23768 standard; DNA; 942 BP.
XX
XX AC ADC23768;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE DNA sequence (SeqID 35) encoding a nitrilase enzyme.
XX
XX KW gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;
XX KW enantiomer; chiral medicine.
XX
XX OS Unidentified.
XX
XX PN WO2003000840-A2.
XX
XX PD 03-JAN-2003.
XX
XX PF 15-MAY-2002; 2002WO-US015983.
XX
XX PR 21-JUN-2001; 2001US-0300189P.
XX PR 30-JUL-2001; 2001US-030906P.
XX PR 22-JAN-2002; 2002US-0351336P.
XX
XX PA (DIVE-) DIVERSA CORP.
XX PA (MADD-) MADDEN D.
XX

PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;
PI Short JM, Burk M;
XX
XX WPI; 2003-201417/19.
XX P-PSDB; ADC23769.
XX
XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl
PT lactic acid derivative and for producing pharmaceutical composition, and
PT food additive.
XX
XX Claim 1; SEQ ID NO 35; 560pp; English.
XX
XX This invention relates to nitrilases and the nucleic acids that encode
CC these enzymes thereof. Specifically, it refers to polypeptides that
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse
CC nitriles or cyanohydrins into their corresponding carboxylic acids and
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in
CC the synthesis of enantiomerically pure aromatic and aliphatic amino
CC acids, as well as hydroxy acids, which are important for the development
CC of chiral medicines. Furthermore, the present invention describes
CC nitrilases, isolated from mesophilic microorganisms, that have improved
CC activity and stability at increased pH and temperature. They are also
CC inexpensive, efficient catalysts, have broad substrate specificity and
CC are capable of chiral differentiation. This polynucleotide is a DNA
CC sequence that encodes a nitrilase enzyme of the invention.
XX
XX SQ Sequence 942 BP; 220 A; 221 C; 275 G; 226 T; 0 U; 0 Other;
Query Match 84.2%; Score 16; DB 9; Length 942;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGACGCCGCCGATG 16
DB 384 GTCGACGCCGCCGATG 369
RESULT 20
AAA53972
ID AAA53972 standard; DNA; 1821 BP.
XX
XX AC AAA53972;
XX
XX DT 08-FEB-2001 (first entry)
XX
XX DE M. tuberculosis Mtb-67.2 antigen.
XX
XX KW Mycobacterium tuberculosis; antigen; Mtb-91; Mtb-67.2; APC;
XX KW antigen presenting cell; serodiagnosis; detection;
XX KW human immunodeficiency virus; HIV; acquired immune deficiency syndrome;
XX KW AIDS; ds.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PH Key Location/Qualifiers
XX FT CDS 1..1821
XX FT /*tag= a
XX FT /product= "Mtb-67.2 antigen"
XX
XX PN WO2000055194-A2.
XX
XX PD 21-SEP-2000.
XX
XX PF 17-MAR-2000; 2000WO-US007196.
XX
XX PR 18-MAR-1999; 99US-00272975.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Hendrickson RC, Lodes MJ, Houghton RL;
XX WPI; 2000-638180/61.
XX

CC and quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. The
CC polypeptides encoded by (I) may be used as antigens in the production of
CC antibodies specific for polymorphic polypeptides. The antibodies may also
CC be used to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of polymorphic
CC polypeptides in samples

XX Sequence 51 BP; 6 A; 21 C; 17 G; 7 T; 0 U; 0 Other;
SQ Query Match 83.2%; Score 15.8; DB 4; Length 51;
Best Local Similarity 89.5%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
DB 5 GCGGACGCCGCCGATGCC 23

RESULT 23
AAI79702
ID AAI79702 standard; DNA; 51 BP.
XX AC AAI79702;
XX DT 09-NOV-2001 (first entry)
XX DE Human conservative amino acid changing SNP nucleic acid SEQ:6643.
XX KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;
XX KW protein therapy; vaccine; probe; diagnostic assay; detection;
XX KW quantitation; restorative therapy; polymorphic; ds.
XX OS Homo sapiens.
XX PN WO200140521-A2.
XX PD 07-JUN-2001.
XX PF 30-NOV-2000; 2000WO-US032758.
XX PR 30-NOV-1999; 99US-0168138P.
XX PR 29-NOV-2000; 2000US-00726173.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX DR WPI; 2001-356160/37.
XX PT Polymorphic nucleic acid sequences, useful in genetic testing and
XX PT therapy.
XX PS Claim 1; Page 2539; 2653pp; English.

XX AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC AAI73114 to AAI73329 represent peptides related to human polymorphic
CC polynucleotide sequences. The sequences can be used in gene and protein
CC therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides. For
CC example, (I) may be used to treat disorders by rectifying mutations or
CC deletions in a patient's genome that affect the activity of polypeptides
CC by expressing inactive proteins or to supplement the patients own
CC production of polypeptide. Additionally, (I) and its complementary
CC sequences may also be used as DNA probes in diagnostic assays to detect
CC and quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. The
CC polypeptides encoded by (I) may be used as antigens in the production of
CC antibodies specific for polymorphic polypeptides. The antibodies may also
CC be used to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of polymorphic

CC polypeptides in samples
XX Sequence 51 BP; 6 A; 22 C; 14 G; 9 T; 0 U; 0 Other;
SQ Query Match 83.2%; Score 15.8; DB 4; Length 51;
Best Local Similarity 89.5%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
DB 19 GTCGACTCCCGGATGACC 37

RESULT 24
ACA42539/c
ID ACA42539 standard; DNA; 477 BP.
XX AC ACA42539;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #24196.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Pseudomonas aeruginosa.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR P-PSDB; ABU38669.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 30409; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway;
CC (8) identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 477 BP; 81 A; 170 C; 150 G; 76 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 7; Length 477;
 Best Local Similarity 89.5%; Pred. No. 1.3e-03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19

Db 114 GTCGACGCGCGCGATGACC 96

RESULT 25

AAC75478
 ID AAC75478 standard; cDNA; 598 BP.

AC AAC75478;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1033 polynucleotide sequence SEQ ID NO:2065.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipariatic; antiparkinsonian; nontropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

OS Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.

XX 31-MAR-1999; 99US-0127607P.

XX 02-APR-1999; 99US-0127636P.

XX 05-APR-1999; 99US-0127728P.

XX 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX P-PSDB; AAB41269.

XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.

PS Claim 5; Page 1559-1560; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipariatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 598 BP; 99 A; 198 C; 172 G; 129 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 3; Length 598;

Best Local Similarity 89.5%; Pred. No. 1.2e-03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19

Db 25 GTCGACGCGCGCGATGACC 43

RESULT 26

ABN21241
 ID ABN21241 standard; cDNA; 598 BP.

AC ABN21241;

XX 24-JUN-2002 (first entry)

XX Human ORFX polynucleotide sequence SEQ ID NO:10959.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis; gene; ss.

OS Homo sapiens.

XX WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

XX 30-MAY-2000; 2000US-0206132P.

XX 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX P-PSDB; ABP05489.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders.
 XX Disclosure; SEQ ID NO 10959; 1037pp; English.
 PS
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 598 BP; 99 A; 198 C; 172 G; 129 T; 0 U; 0 Other;
 Query Match 83.2%; Score 15.8; DB 6; Length 598;
 Best Local Similarity 89.5%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GTCGACGCCGCCGATGACC 19
 D5 25 GCCGACGCCGCCGATGCC 43
 RESULT 27
 AAF60781/c
 ID AAF60781 standard; DNA; 609 BP.
 AC AAF60781;
 XX
 XX 04-MAY-2001 (first entry)
 DT
 DE Pseudomonas sp ABC transporter encoding DNA ORF09149.
 XX
 XX ABC transporter; transgenic plant; plant growth promoter; plant yield;
 KW plant development; xenobiotic; soil pollution; inorganic nutrient;
 KW soil quality; plant protection; ds.
 XX
 OS Pseudomonas sp.
 XX
 XX DE19934719-A1.
 PN
 XX
 XX 25-JAN-2001.
 PD
 XX 23-JUL-1999; 99DE-01034719.
 PF
 XX 23-JUL-1999; 99DE-01034719.
 PR
 XX (TIGR-) TIGR INST GENOMIC RES.
 PA (QUIA-) QUIAGEN GMBH.
 PA (GBF-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX
 XX Fraser CM, Venter C, Tuemmler B, Hohnselt J, Duysterhoeft A;
 PI Hilbert H, Timmis K, Moore E, Straetz M, Helm S;
 XX

DR WPI; 2001-160595/17.
 XX New DNA encoding ABC transporters, useful e.g. for promoting growth of
 PT plants, particularly legumes on soil polluted with xenobiotics which are
 PT normally harmful to the plants.
 XX
 PS Claim 3a; Page 34; 82pp; German.
 XX
 CC This invention describes novel DNA sequences (I) that express products
 CC having the biological function of ABC transporters. The invention also
 CC describes (1) recombinant expression vectors containing (I); (2)
 CC prokaryotic and eukaryotic cells transformed or transfected with (I) or
 CC the vector of (a); (3) production of ABC transporters by culturing cells
 CC of (b); (4) expression products (II) of (I), and synthetic proteins or
 CC peptides with the same sequences; (5) mono- or poly-clonal antibodies
 CC (Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
 CC and (7) transgenic plants that contain cells of (b), (I), and their
 CC fragments, are useful for expression of recombinant ABC transporters, and
 CC as probes and primers for detection, isolation and amplification of full-
 CC length cDNA sequences. (I) can also be used to produce transgenic plants.
 CC Cells transformed with (I), recombinant ABC transporters and synthetic
 CC peptides or proteins with the same activities are used to promote growth,
 CC development and yield of plants, particularly Leguminosae, especially
 CC where growing on soil polluted by xenobiotics injurious to plants. Host
 CC cells that express ABC transporters take up and eliminate pollutants, and
 CC release inorganic nutrients, endogenous or modified metabolites and
 CC endogenous or heterologous proteins, i.e. they improve soil quality and
 CC protect plants against pollutants from the soil
 XX
 SQ Sequence 609 BP; 72 A; 202 C; 198 G; 132 T; 0 U; 5 Other;
 Query Match 83.2%; Score 15.8; DB 4; Length 609;
 Best Local Similarity 89.5%; Pred. No. 1.2e+03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GTCGACGCCGCCGATGACC 19
 D5 243 GTCGATGCCGCCGATGATC 225
 RESULT 28
 ADC75920/c
 ID ADC75920 standard; DNA; 620 BP.
 AC ADC75920;
 XX
 XX 01-JAN-2004 (first entry)
 DT
 DE DNA homologous to phytopathogen resistance-related cDNA - SEQ ID 844.
 XX
 XX rice; yeast; poppy; plant; disease resistance; anti-fungal;
 KW phytopathogen; gene shuffling; ds.
 XX
 OS Unidentified.
 XX
 XX WO2003020905-A2.
 PN
 XX
 XX 13-MAR-2003.
 PD
 XX 30-AUG-2002; 2002WO-US027883.
 PF
 XX 31-AUG-2001; 2001US-0316392P.
 PR
 XX (DOWC) DOW CHEM CO.
 PA
 XX Shukla V, Butler H, Larrinua I, Reddy AS;
 PI
 XX WPI; 2003-290185/28.
 DR
 XX Novel isolated nucleic acid derived from Nicotiana benthamiana, Oryzae
 PT sativa, Saccharomyces cerevisiae, Trichoderma harzianum and Papaver
 PT rhoeas, useful for conferring disease resistance in plants.
 XX

PS Claim 1; SEQ ID NO 844; 617pp; English.

XX The invention relates to a novel isolated nucleic acid derived from

CC Nicotiana benthamiana, Oryza sativa (rice), Saccharomyces cerevisiae

CC (yeast), Trichoderma harzianum (Hypocrea lixii) and Papaver rhoeas

CC (poppy) and a sequence that hybridises to them under conditions of low

CC stringency, where expression of the nucleic acid in a plant results in a

CC disease resistance phenotype. The polynucleotides of the invention

CC demonstrate anti-fungal activity and may be useful in conferring disease

CC resistance in a plant against phytopathogen such as Aspergillus flavus,

CC Gibberella fujikuroi and Gibberella zeae. Furthermore, the

CC polynucleotides may be useful to retrieve unknown sequences and in gene

CC shuffling or sexual PCR procedures. The current sequence is that of the

CC DNA of the invention which is homologous to that of the phytopathogen

CC resistance-related contig cDNAs.

XX SQ Sequence 620 BP; 101 A; 227 C; 192 G; 100 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 9; Length 620;

Best Local Similarity 89.5%; Pred. NO. 1.2e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGCGCGCGCGATGACC 19

Db 331 GACGAGCGCGCGATGACC 313

RESULT 29

ACA42466/c

ID ACA42466 standard; DNA; 738 BP.

XX ACA42466;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #24123.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX Pseudomonas aeruginosa.

OS WO20027183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI WPI; 2003-029926/02.

DR P-PSDB; ABU38596.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 30336; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 738 BP; 131 A; 224 C; 245 G; 138 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 7; Length 738;

Best Local Similarity 89.5%; Pred. NO. 1.2e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGCGAGCGCGCGATGACC 19

Db 186 GTGCGCGCGCGATGACC 168

RESULT 30

ADC98771/c

ID ADC98771 standard; DNA; 738 BP.

XX ADC98771;

XX 01-JAN-2004 (first entry)

DE Pseudomonas aeruginosa pyrH DNA - SEQ ID 64.

XX adenine phosphoribosyltransferase; apt; uridylylate kinase; pyrH;

KW guanylate kinase; gmk; antibacterial; vaccine; bronchial infection;

KX sinusitis; meningitis; food preservative; disinfectant; ds; gene.

XX Pseudomonas aeruginosa.

OS WO2003035858-A2.

PN 01-MAY-2003.

XX 25-OCT-2002; 2002WO-CA001613.

XX 25-OCT-2001; 2001US-0337625P.

PR 26-OCT-2001; 2001US-0340534P.

PR 18-DEC-2001; 2001US-0341639P.

PR 18-DEC-2001; 2001US-0341825P.

PR 19-DEC-2001; 2001US-0342004P.

PR 20-DEC-2001; 2001US-0342559P.

XX (AFFI-) AFFINIUM PHARM INC.

XX Edwards A, Dharamsi A, Vedadi M, Alam MZ, Awrey D, Beattie B;

PI Canadian V, Domagala M, Houston S, Li Q, Mansoury K, Necakov S;

PI Ng I, Pinder B, Sheldrick B, Wrezel O;

DR WPI; 2003-421420/39.
 DR P-PSDB; ADC98772.
 XX
 PT Novel recombinant polypeptide useful for designing a modulator for the
 PT prevention or treatment of Streptococcus aureus related disease or
 PT disorder.
 XX
 PS Claim 136; SEQ ID NO 64; 252pp; English.
 XX
 CC The invention relates to a novel composition comprising an isolated,
 CC recombinant polypeptide. The polypeptide may comprise Staphylococcus
 CC aureus or Streptococcus pneumoniae adenine phosphoribosyltransferase
 CC (apt), uridylylate kinase (pyrH) isolated from Staphylococcus aureus,
 CC Streptococcus pneumoniae or Pseudomonas aeruginosa or S. pneumoniae
 CC guanylate kinase (gmk). The polypeptides of the invention demonstrate
 CC antibacterial activity whilst the composition may be useful for designing
 CC a modulator facilitating the prevention or treatment of an S. aureus or
 CC S. pneumoniae related disease or disorder, such as bronchial infection,
 CC sinusitis and meningitis. Furthermore, the composition may be utilised to
 CC generate a vaccine for protection against invasion of bacteria, as well
 CC as to develop food preservatives and surface disinfectants. The current
 CC sequence is that of the P. aeruginosa pyrH DNA of the invention.
 XX
 SQ Sequence 738 BP; 131 A; 225 C; 245 G; 137 T; 0 U; 0 Other;
 Query Match 83.2%; Score 15.8; DB 9; Length 738;
 Best Local Similarity 89.5%; Pred. No. 1.2e-03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GTCGACGCCGCGATGACC 19
 Db 186 GTTGGCGCGCGATGACC 168

Search completed: June 20, 2004, 10:17:49
 Job time : 122.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 887.973 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19
Sequence: 1 gtcgacgcgcgcgatgacc 19

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27533289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estnu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_estc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_estc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vit.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_pig.*
- 27: em_gss_vil.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match Length | ID | Description |
|------------|-------|--------------|-----|--------------------|
| 1 | 19 | 100.0 | 721 | BI099738 602884915 |
| 2 | 18 | 94.7 | 476 | BG153958 188 t1n01 |
| 3 | 17.4 | 91.6 | 182 | AL753282 Arabidops |
| 4 | 17.4 | 91.6 | 200 | AU181058 AU181058 |

| | | | | | | | | | | | | | | | |
|-----|------|------|-----|----|-----------|------------|-----|------|------|------|----|----------|----------|----------|-----------|
| 78 | 16.4 | 86.3 | 652 | 10 | BE357618 | DGL 21 D0 | 151 | 16.4 | 86.3 | 959 | 14 | CK254221 | CK254221 | CK254221 | EST737558 |
| 79 | 16.4 | 86.3 | 654 | 28 | BH928584 | BACPP27-N | 152 | 16.4 | 86.3 | 963 | 14 | CK244036 | CK244036 | CK244036 | EST727673 |
| 80 | 16.4 | 86.3 | 656 | 14 | CF627049 | zmtws05 0 | 153 | 16.4 | 86.3 | 967 | 14 | CK256446 | CK256446 | CK256446 | EST740083 |
| 81 | 16.4 | 86.3 | 663 | 13 | BQ462905 | HI02G03F | 154 | 16.4 | 86.3 | 975 | 14 | CK272783 | CK272783 | CK272783 | EST718861 |
| 82 | 16.4 | 86.3 | 664 | 29 | CG354182 | OGXIC39TH | 155 | 16.4 | 86.3 | 979 | 14 | CK249986 | CK249986 | CK249986 | EST733623 |
| 83 | 16.4 | 86.3 | 665 | 29 | CG258424 | OGXIC39TH | 156 | 16.4 | 86.3 | 983 | 14 | CK251027 | CK251027 | CK251027 | EST734664 |
| 84 | 16.4 | 86.3 | 669 | 14 | CD883845 | OGXIC39TH | 157 | 16.4 | 86.3 | 982 | 14 | CK254225 | CK254225 | CK254225 | EST737862 |
| 85 | 16.4 | 86.3 | 670 | 9 | AU070263 | AU070263 | 158 | 16.4 | 86.3 | 1413 | 11 | AY103665 | AY103665 | AY103665 | zmtws05 0 |
| 86 | 16.4 | 86.3 | 671 | 14 | CF631174 | zmtws05 0 | 159 | 16.4 | 86.3 | 207 | 12 | BM733596 | BM733596 | BM733596 | zmtws05 0 |
| 87 | 16.4 | 86.3 | 677 | 14 | CA290105 | SCAGFL801 | 160 | 16.4 | 86.3 | 312 | 29 | FR005620 | FR005620 | FR005620 | zmtws05 0 |
| 88 | 16.4 | 86.3 | 677 | 14 | CF628006 | zmtws05 0 | 161 | 16.4 | 86.3 | 319 | 9 | AA000418 | AA000418 | AA000418 | zmtws05 0 |
| 89 | 16.4 | 86.3 | 683 | 14 | CK253866 | EST737503 | 162 | 16.4 | 86.3 | 326 | 13 | C25993 | C25993 | C25993 | zmtws05 0 |
| 90 | 16.4 | 86.3 | 684 | 29 | CG617972 | OGWBYA3TV | 163 | 16.4 | 86.3 | 366 | 9 | AU065167 | AU065167 | AU065167 | zmtws05 0 |
| 91 | 16.4 | 86.3 | 685 | 14 | CA171503 | SCSBSB105 | 164 | 16.4 | 86.3 | 388 | 14 | CB639395 | CB639395 | CB639395 | zmtws05 0 |
| 92 | 16.4 | 86.3 | 685 | 14 | CF627858 | zmtws05 0 | 165 | 16.4 | 86.3 | 501 | 14 | CF153460 | CF153460 | CF153460 | zmtws05 0 |
| 93 | 16.4 | 86.3 | 686 | 13 | CA102475 | SCRGHR106 | 166 | 16.4 | 86.3 | 510 | 14 | CD680055 | CD680055 | CD680055 | zmtws05 0 |
| 94 | 16.4 | 86.3 | 688 | 14 | CA484842 | WH4311 C | 167 | 16.4 | 86.3 | 528 | 28 | AQ850161 | AQ850161 | AQ850161 | zmtws05 0 |
| 95 | 16.4 | 86.3 | 688 | 14 | CA756082 | BR0300340 | 168 | 16.4 | 86.3 | 529 | 9 | AU239902 | AU239902 | AU239902 | zmtws05 0 |
| 96 | 16.4 | 86.3 | 693 | 10 | AW565610 | LGI 346 E | 169 | 16.4 | 86.3 | 536 | 13 | BU999236 | BU999236 | BU999236 | zmtws05 0 |
| 97 | 16.4 | 86.3 | 694 | 14 | CA201745 | SCRGHR106 | 170 | 16.4 | 86.3 | 538 | 12 | BM735479 | BM735479 | BM735479 | zmtws05 0 |
| 98 | 16.4 | 86.3 | 694 | 28 | B2962628 | PURIE86TB | 171 | 16.4 | 86.3 | 547 | 28 | AZ935445 | AZ935445 | AZ935445 | zmtws05 0 |
| 99 | 16.4 | 86.3 | 699 | 12 | BI406624 | 172C11 Ma | 172 | 16.4 | 86.3 | 550 | 14 | CF939510 | CF939510 | CF939510 | zmtws05 0 |
| 100 | 16.4 | 86.3 | 699 | 14 | CD218960 | PG1n.pk0 | 173 | 16.4 | 86.3 | 575 | 14 | CF959315 | CF959315 | CF959315 | zmtws05 0 |
| 101 | 16.4 | 86.3 | 702 | 9 | AU055864 | AU055864 | 174 | 16.4 | 86.3 | 579 | 14 | CF940242 | CF940242 | CF940242 | zmtws05 0 |
| 102 | 16.4 | 86.3 | 704 | 14 | CA176030 | SCULST102 | 175 | 16.4 | 86.3 | 581 | 14 | CF597776 | CF597776 | CF597776 | zmtws05 0 |
| 103 | 16.4 | 86.3 | 706 | 14 | CD863279 | AZ01.106E | 176 | 16.4 | 86.3 | 581 | 28 | AZ934268 | AZ934268 | AZ934268 | zmtws05 0 |
| 104 | 16.4 | 86.3 | 706 | 14 | CK259264 | EST742901 | 177 | 16.4 | 86.3 | 585 | 14 | CF942075 | CF942075 | CF942075 | zmtws05 0 |
| 105 | 16.4 | 86.3 | 712 | 28 | AZ130497 | OSUNBB010 | 178 | 16.4 | 86.3 | 585 | 14 | CF659907 | CF659907 | CF659907 | zmtws05 0 |
| 106 | 16.4 | 86.3 | 714 | 10 | BE357648 | DGL 21 D0 | 179 | 16.4 | 86.3 | 615 | 13 | CA137620 | CA137620 | CA137620 | zmtws05 0 |
| 107 | 16.4 | 86.3 | 714 | 13 | CA072793 | SCPEAM102 | 180 | 16.4 | 86.3 | 619 | 14 | CF421897 | CF421897 | CF421897 | zmtws05 0 |
| 108 | 16.4 | 86.3 | 717 | 14 | CK243563 | EST727200 | 181 | 16.4 | 86.3 | 663 | 14 | CF422251 | CF422251 | CF422251 | zmtws05 0 |
| 109 | 16.4 | 86.3 | 719 | 14 | CB678634 | OSUNEB16P | 182 | 16.4 | 86.3 | 663 | 14 | CF422251 | CF422251 | CF422251 | zmtws05 0 |
| 110 | 16.4 | 86.3 | 724 | 14 | CD425969 | SAI 15 F1 | 183 | 16.4 | 86.3 | 665 | 29 | CG759814 | CG759814 | CG759814 | zmtws05 0 |
| 111 | 16.4 | 86.3 | 729 | 14 | CD881998 | FL 104F10 | 184 | 16.4 | 86.3 | 706 | 14 | CF874693 | CF874693 | CF874693 | zmtws05 0 |
| 112 | 16.4 | 86.3 | 730 | 13 | CA152520 | SCJFR2202 | 185 | 16.4 | 86.3 | 740 | 28 | B2530003 | B2530003 | B2530003 | zmtws05 0 |
| 113 | 16.4 | 86.3 | 737 | 14 | CD099820 | G468.113 K | 186 | 16.4 | 86.3 | 752 | 29 | CG620372 | CG620372 | CG620372 | zmtws05 0 |
| 114 | 16.4 | 86.3 | 743 | 14 | CD430205 | ETH1.17 G | 187 | 16.4 | 86.3 | 751 | 14 | CB896633 | CB896633 | CB896633 | zmtws05 0 |
| 115 | 16.4 | 86.3 | 747 | 13 | BQ803135 | WH2834 B | 188 | 16.4 | 86.3 | 762 | 12 | BG321279 | BG321279 | BG321279 | zmtws05 0 |
| 116 | 16.4 | 86.3 | 755 | 29 | CA2187812 | PU6P40TB | 189 | 16.4 | 86.3 | 771 | 29 | CG288032 | CG288032 | CG288032 | zmtws05 0 |
| 117 | 16.4 | 86.3 | 778 | 14 | CA2187812 | PU6P40TB | 190 | 16.4 | 86.3 | 772 | 14 | CB903186 | CB903186 | CB903186 | zmtws05 0 |
| 118 | 16.4 | 86.3 | 782 | 14 | CA221025 | SCRUFL402 | 191 | 16.4 | 86.3 | 772 | 14 | CB903186 | CB903186 | CB903186 | zmtws05 0 |
| 119 | 16.4 | 86.3 | 783 | 14 | CB673195 | OSUNEB07J | 192 | 16.4 | 86.3 | 796 | 28 | B2810952 | B2810952 | B2810952 | zmtws05 0 |
| 120 | 16.4 | 86.3 | 795 | 14 | CK252664 | EST736301 | 193 | 16.4 | 86.3 | 812 | 14 | CB649283 | CB649283 | CB649283 | zmtws05 0 |
| 121 | 16.4 | 86.3 | 799 | 28 | AQ576888 | nmxb0089P | 194 | 16.4 | 86.3 | 817 | 29 | CG137004 | CG137004 | CG137004 | zmtws05 0 |
| 122 | 16.4 | 86.3 | 805 | 13 | BQ838243 | WH2908 B | 195 | 16.4 | 86.3 | 900 | 29 | CG222100 | CG222100 | CG222100 | zmtws05 0 |
| 123 | 16.4 | 86.3 | 807 | 14 | CA2187812 | PU6P40TB | 196 | 16.4 | 86.3 | 908 | 14 | CF885408 | CF885408 | CF885408 | zmtws05 0 |
| 124 | 16.4 | 86.3 | 810 | 14 | CA2187812 | PU6P40TB | 197 | 16.4 | 86.3 | 916 | 29 | CG112683 | CG112683 | CG112683 | zmtws05 0 |
| 125 | 16.4 | 86.3 | 813 | 29 | CG731056 | OGWDE61TV | 198 | 16.4 | 86.3 | 940 | 29 | CG162917 | CG162917 | CG162917 | zmtws05 0 |
| 126 | 16.4 | 86.3 | 819 | 29 | CG731056 | OGWDE61TV | 199 | 16.4 | 86.3 | 958 | 29 | CG117701 | CG117701 | CG117701 | zmtws05 0 |
| 127 | 16.4 | 86.3 | 820 | 13 | CA090915 | SCUTAM200 | 200 | 15.8 | 83.2 | 118 | 12 | B2942373 | B2942373 | B2942373 | zmtws05 0 |
| 128 | 16.4 | 86.3 | 828 | 14 | CK254960 | EST738597 | 201 | 15.8 | 83.2 | 142 | 12 | CB960816 | CB960816 | CB960816 | zmtws05 0 |
| 129 | 16.4 | 86.3 | 828 | 28 | BZ565381 | PGS2.164 | 202 | 15.8 | 83.2 | 144 | 11 | CNS0901F | CNS0901F | CNS0901F | zmtws05 0 |
| 130 | 16.4 | 86.3 | 829 | 29 | CG668217 | OSUNEB16P | 203 | 15.8 | 83.2 | 165 | 14 | CB826026 | CB826026 | CB826026 | zmtws05 0 |
| 131 | 16.4 | 86.3 | 838 | 14 | CK251853 | EST735490 | 204 | 15.8 | 83.2 | 167 | 12 | BU188789 | BU188789 | BU188789 | zmtws05 0 |
| 132 | 16.4 | 86.3 | 861 | 14 | CB643663 | OSUNEB07J | 205 | 15.8 | 83.2 | 180 | 6 | AL813023 | AL813023 | AL813023 | zmtws05 0 |
| 133 | 16.4 | 86.3 | 867 | 29 | CG203456 | PUFT23TB | 206 | 15.8 | 83.2 | 180 | 11 | CNS0935P | CNS0935P | CNS0935P | zmtws05 0 |
| 134 | 16.4 | 86.3 | 870 | 14 | CK253865 | EST737502 | 207 | 15.8 | 83.2 | 181 | 10 | B400192 | B400192 | B400192 | zmtws05 0 |
| 135 | 16.4 | 86.3 | 872 | 29 | CG437693 | OGVGB08TH | 208 | 15.8 | 83.2 | 189 | 10 | B594948 | B594948 | B594948 | zmtws05 0 |
| 136 | 16.4 | 86.3 | 880 | 28 | BZ636180 | OSUNEB16P | 209 | 15.8 | 83.2 | 226 | 9 | AV642285 | AV642285 | AV642285 | zmtws05 0 |
| 137 | 16.4 | 86.3 | 881 | 28 | CC192620 | OSUNEB16P | 210 | 15.8 | 83.2 | 227 | 12 | BM134902 | BM134902 | BM134902 | zmtws05 0 |
| 138 | 16.4 | 86.3 | 889 | 10 | BF127234 | 1191d11.9 | 211 | 15.8 | 83.2 | 236 | 28 | BZ328816 | BZ328816 | BZ328816 | zmtws05 0 |
| 139 | 16.4 | 86.3 | 891 | 14 | CK247284 | 601650891 | 212 | 15.8 | 83.2 | 240 | 9 | AJ474212 | AJ474212 | AJ474212 | zmtws05 0 |
| 140 | 16.4 | 86.3 | 903 | 14 | CK244035 | EST727672 | 213 | 15.8 | 83.2 | 246 | 10 | BE500426 | BE500426 | BE500426 | zmtws05 0 |
| 141 | 16.4 | 86.3 | 903 | 14 | CK249263 | EST742900 | 214 | 15.8 | 83.2 | 250 | 9 | AV626586 | AV626586 | AV626586 | zmtws05 0 |
| 142 | 16.4 | 86.3 | 910 | 14 | CK259263 | EST727672 | 215 | 15.8 | 83.2 | 257 | 13 | CA024181 | CA024181 | CA024181 | zmtws05 0 |
| 143 | 16.4 | 86.3 | 912 | 14 | CK249734 | EST734593 | 216 | 15.8 | 83.2 | 261 | 9 | AV626586 | AV626586 | AV626586 | zmtws05 0 |
| 144 | 16.4 | 86.3 | 913 | 14 | CK249734 | EST734593 | 217 | 15.8 | 83.2 | 261 | 11 | CNS08ECC | CNS08ECC | CNS08ECC | zmtws05 0 |
| 145 | 16.4 | 86.3 | 917 | 28 | CC375337 | PUHRS74TB | 218 | 15.8 | 83.2 | 262 | 29 | CC734842 | CC734842 | CC734842 | zmtws05 0 |
| 146 | 16.4 | 86.3 | 922 | 13 | BQ674430 | AGENCOURT | 219 | 15.8 | 83.2 | 267 | 11 | CNS09BPA | CNS09BPA | CNS09BPA | zmtws05 0 |
| 147 | 16.4 | 86.3 | 929 | 14 | CK249734 | EST734593 | 220 | 15.8 | 83.2 | 270 | 9 | AV435832 | AV435832 | AV435832 | zmtws05 0 |
| 148 | 16.4 | 86.3 | 939 | 11 | AY104963 | zmtws05 0 | 221 | 15.8 | 83.2 | 275 | 14 | D24235 | D24235 | D24235 | zmtws05 0 |
| 149 | 16.4 | 86.3 | 950 | 28 | BZ791435 | PUFFK25TB | 222 | 15.8 | 83.2 | 277 | 14 | CA659664 | CA659664 | CA659664 | zmtws05 0 |
| 150 | 16.4 | 86.3 | 957 | 14 | CK250012 | EST733649 | 223 | 15.8 | 83.2 | 279 | 11 | CNS08X7T | CNS08X7T | CNS08X7T | zmtws05 0 |

C 224 15.8 83.2 280 14 CK145179 3530.1.16
C 225 15.8 83.2 283 11 CQ509611
C 226 15.8 83.2 287 12 BM375241
C 227 15.8 83.2 290 12 BM642572
C 228 15.8 83.2 290 13 BQ760616
C 229 15.8 83.2 291 18 AZ578266
C 230 15.8 83.2 298 13 BX767139
C 231 15.8 83.2 299 12 BML38699
C 232 15.8 83.2 300 9 AJ436424
C 233 15.8 83.2 301 11 CQ509KLO
C 234 15.8 83.2 304 11 CQ509F9S
C 235 15.8 83.2 314 11 CQ509KY3
C 236 15.8 83.2 319 14 CA670677
C 237 15.8 83.2 318 13 CA019691
C 238 15.8 83.2 325 14 D24505
C 239 15.8 83.2 328 13 CA026320
C 240 15.8 83.2 329 11 CQ50907Q
C 241 15.8 83.2 331 9 AJ280053
C 242 15.8 83.2 332 11 CQ508GAP
C 243 15.8 83.2 334 12 BM259076
C 244 15.8 83.2 337 9 AV629675
C 245 15.8 83.2 337 12 BG607614
C 246 15.8 83.2 340 11 CQ508CJX
C 247 15.8 83.2 341 11 CQ508ZJX
C 248 15.8 83.2 343 13 BX767371
C 249 15.8 83.2 345 10 BE637925
C 250 15.8 83.2 347 28 BH583552
C 251 15.8 83.2 351 11 CQ5098F5
C 252 15.8 83.2 353 28 BH639855
C 253 15.8 83.2 354 9 AV630085
C 254 15.8 83.2 354 14 CA643930
C 255 15.8 83.2 357 11 CQ508L7G
C 256 15.8 83.2 357 28 BH678798
C 257 15.8 83.2 358 9 AV630718
C 258 15.8 83.2 358 11 CQ508G9K
C 259 15.8 83.2 359 11 CQ509I21
C 260 15.8 83.2 360 9 AJ474213
C 261 15.8 83.2 360 9 AV629992
C 262 15.8 83.2 361 9 AJ281144
C 263 15.8 83.2 362 29 CQ5010F6
C 264 15.8 83.2 364 14 CD163241
C 265 15.8 83.2 366 11 CQ50900B
C 266 15.8 83.2 371 11 CQ508091
C 267 15.8 83.2 375 13 BX766047
C 268 15.8 83.2 378 28 AQ902476
C 269 15.8 83.2 383 9 AU165848
C 270 15.8 83.2 384 13 AC843713
C 271 15.8 83.2 384 28 AC843713
C 272 15.8 83.2 386 9 AU162306
C 273 15.8 83.2 387 12 BM368351
C 274 15.8 83.2 390 12 BM644751
C 275 15.8 83.2 390 14 CB816855
C 276 15.8 83.2 390 29 CG765065
C 277 15.8 83.2 392 12 BJ201699
C 278 15.8 83.2 393 13 BQ767691
C 279 15.8 83.2 398 12 BG544751
C 280 15.8 83.2 400 12 BX598528
C 281 15.8 83.2 400 13 BX767307
C 282 15.8 83.2 400 28 AZ919944
C 283 15.8 83.2 402 9 AJ284757
C 284 15.8 83.2 403 13 BQ755135
C 285 15.8 83.2 406 13 BX601762
C 286 15.8 83.2 406 13 BX767448
C 287 15.8 83.2 406 13 BX769027
C 288 15.8 83.2 407 13 BX766546
C 289 15.8 83.2 410 13 BX768792
C 290 15.8 83.2 411 11 CQ5096RR
C 291 15.8 83.2 411 14 CB212396
C 292 15.8 83.2 412 12 BG159499
C 293 15.8 83.2 412 13 CA002701
C 294 15.8 83.2 412 13 CA002765
C 295 15.8 83.2 416 11 CQ509474
C 296 15.8 83.2 417 11 CQ509GGQ

CK145179 3530.1.16
BX043889 Single re
BM375241 EBem06 SQ
BM642572 170006590
BQ760616 EBro03 SQ
AZ578266 20h04 Shro
BX767139 BX767139
BML38699 WHE0496 H
AJ436424 AJ436424
BX062776 Single re
BX055868 Single re
BX063223 Single re
CA670677 wlsul.pk0
CA019691 HV12M17r
CA026320 R255J04r
BX036354 Single re
AJ280053 4A3A-AAK-
BM259076 TaF905.B3
AV629675 AV629675
BG607614 WHE2481.B
BX006102 Single re
BX035497 Single re
BX767371 BX767371
BH583552 BQGTZ73TF
BE637925 WHE1755-1
BX046989 Single re
BH639855 10C8032D1
AV630085 AV630085
CA643930 wreln.pk0
BX017696 Single re
BH678798 BOHSQ58TR
AV630718 AV630718
BX010500 Single re
BX059477 Single re
AJ474213 AJ474213
AV629992 AV629992
AJ281144 4A3A-AB-
AL153442 Anopheles
CD163241 ML1-0082U
BX068127 Single re
BX020849 Single re
BX766047 BX766047
AQ902476 LNAJFV1.1
AU165848 AU165848
BX767488 BX767488
AC843713 nbxb0024P
AU162306 AU162306
BM368351 EBed01 SQ
BM644751 170006590
CB816855 Gbl12pz.r
CG765065 ZMMBSc020
BJ201699 BJ201699
BQ767691 EBro08 SQ
BG544751 E2720 Chi
BM598528 170006876
BX767307 BX767307
AZ919944 1006017D0
AJ284757 4A3B-AB-
BQ755135 EBed02 SQ
BX601762 BX601762
BX767448 BX767448
BX769027 BX769027
BX766546 BX766546
BX768792 BX768792
EX044851 Single re
CB212396 OML02676
BG159499 OV2.5.B05
CA002701 HS08H1r
CA002765 HS08H10r
EX041516 Single re
EX057414 Single re

C 297 15.8 83.2 417 12 BM584771
C 298 15.8 83.2 419 28 AQ847623
C 299 15.8 83.2 422 12 BM617375
C 300 15.8 83.2 422 29 AG211681

ALIGNMENTS

RESULT 1
BI099738 721 bp mRNA linear EST 26-JUN-2001
LOCUS 602884916F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5040602
DEFINITION 5', mRNA sequence.
ACCESSION BI099738
VERSION BI099738.1 GI:14550631
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 721)
AUTHORS NIH-WGAC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1111 row: 0 column: 03
High quality sequence stop: 589.

FEATURES
source
1..721
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5040602"
/lab_host="PH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Kid14"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"

ORIGIN
Query Match 100.0%; Score 19; DB 12; Length 721;
Best Local Similarity 100.0%; Pred. No. 8e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGACGCCGCCGATGACC 19
|||||
DB 549 GTCGACGCCGCCGATGACC 567
|||||

RESULT 2
BI153958 476 bp mRNA linear EST 05-FEB-2001
LOCUS 188 LLN01 Lupinus luteus cDNA, mRNA sequence.
DEFINITION BI153958
ACCESSION BI153958.1 GI:12665988
VERSION BI153958.1
KEYWORDS EST.
SOURCE Lupinus luteus (yellow lupine)
ORGANISM Lupinus luteus
REFERENCE 1 (bases 1 to 476)
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.

REFERENCE 1 (bases 1 to 476)
 AUTHORS Podkowinski, J., Kistiel, A., Grabowska, B. and Legocki, A. B.
 TITLE ESTs from the early nodules library of *Lupinus luteus*
 JOURNAL Unpublished (2000)
 COMMENT Contact: Podkowinski J
 Laboratory of Plant Molecular Biology
 Institute of Bioorganic Chemistry of Polish Academy of Sciences
 Noskowskiego 12/14, 61-704 Poznan, Poland
 Tel: 0048 61 8528919
 Fax: 0048 61 8520532
 Email: jantsp@ibch.poznan.pl
 POLYA=Yes.

FEATURES

source

1. 476
 /organism="Lupinus luteus"
 /mol_type="mRNA"
 /cultivar="ventus"
 /db_xref="taxon:3873"
 /tissue_type="roots with young nodules"
 /dev_stage="roots with immature nodules harvested seven
 days post inoculation with *Bradyrhizobium* sp. WM9
 (lupinus)"
 /lab_host="E. coli strain SOLR"
 /clone_lib="LIN01"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots with developing nodules harvested seven days post
 inoculation with *Bradyrhizobium* sp. WM9 (lupinus). The
 cDNA was directionally ligated into the Uni-ZAP XR vector
 from Stratagene and packaged using Gigapack III Gold
 packaging extracts. Plasmids containing cDNA inserts were
 excised from the recombinant lambda-ZAP phage using
 Ex-Assist helper phage and propagated in SOLR cells."

ORIGIN

Query Match 94.7%; Score 18; DB 12; Length 476;
 Best Local Similarity 100.0%; Pred. No. 1.7e+04; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 0

QY 2 TCGACGCCGCCGATGACC 19
 |||||
 Db 235 TCGACGCCGCCGATGACC 252

RESULT 3
 AL759282/c
 LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-167C03-013362,
 DEFINITION genomic survey sequence.
 ACCESSION AL759282 182 bp DNA linear GSS 18-JUN-2002
 VERSION AL759282.1 GI:21497630
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
 AUTHORS Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.
 and Weisshaar, B.
 TITLE A pipeline for automated high-throughput generation of FSTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines
 JOURNAL Unpublished

REFERENCE 2
 AUTHORS Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
 TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 182)
 AUTHORS Strizhov, N., Rosso, M., Li, Y. and Weisshaar, B.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

COMMENT

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone t20g20.
 The sequences are generated at the MPI for Plant Breeding Research
 in the context of the GABI-Kat project. GABI-Kat is part of the
 German Plant Genomics program designated 'GABI'. Information on
 line availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers
 1. 182
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="GK-167C03-013362"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana
 plants (Ti) which were transformed with the T-DNA from
 vector pAC161. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed"

ORIGIN

Query Match 91.6%; Score 17.4; DB 29; Length 182;
 Best Local Similarity 94.7%; Pred. No. 2.5e+04; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 1

QY 1 GTCGACGCCGCCGATGACC 19
 |||||
 Db 112 GTCGACGCCGCCGATGACC 94

RESULT 4

LOCUS AU181058
 DEFINITION AU181058 Rice shoot Oryza sativa (japonica cultivar-group) cDNA
 clone S6021, mRNA sequence.
 ACCESSION AU181058 200 bp mRNA linear EST 03-APR-2002
 VERSION AU181058.1 GI:13447564
 KEYWORDS EST.

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

1 (bases 1 to 200)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from etiolated shoot (2001)
 Unpublished (2001)
 JOURNAL Contact: Takuji Sasaki
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@abr.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>
 PROJECT = 'RGP'.
 Location/Qualifiers

FEATURES

source

1. 200
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="S6021"
 /dev_stage="Etiolated shoot (8 days old)"
 /clone_lib="Rice shoot"
 /note="Etiolated shoot (8 days old)"

ORIGIN

Total hg bases = 321
 Seq primer: ATTTACCTCTACTAAGGG
 High quality sequence stop: 383.
 Location/Qualifiers
 1. .874

FEATURES

source
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVGMEF0001M19f"
 /tissue_type="Seedling root"
 /lab_host="TJC121"
 /clone_lib="Hordeum vulgare seedling root EST library
 HVCDNA0007 (Etiolated and unstressed)"
 /notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper
 with water, nystatin and cefotaxime in covered
 crystallization dishes. Five-day old seedling roots were
 then harvested, total RNA was prepared, poly(A) RNA was
 purified, one primary unamplified cDNA library was made,
 and 1 million pfu were in vivo excised to give pBluescript
 SK(-) cDNA phagemids. These steps were performed in the IU
 Close laboratory at the University of California,
 Riverside (Choi, Close, Fenton). Phagemids were plated and
 picked at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
 Gates, Rambo, Main). The sequence has been trimmed to
 remove vector sequence and contains a minimum of 100 bases
 of phred value 20 or above. For more details on library
 preparation and sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
 this clone see <http://www.genome.clemson.edu/orders> Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley Genetics Newsletter 31:29-30.
 (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 874;
 Best Local Similarity 94.7%; Pred. No. 2.8e+04;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGACC 19
 |||||
 Db 180 GTCGACGCCGCGGTGACC 162

RESULT 8

CG172122
 LOCUS
 DEFINITION
 CG172122 881 bp DNA linear GSS 21-AUG-2003
 PUJAX68TB ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTa0626Li6,
 genomic survey sequence.

ACCESSION
 CG172122
 VERSION
 CG172122.1 GI:34062920
 KEYWORDS
 GSS.

ORGANISM

Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 881)

REFERENCE
 AUTHORS
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

TITLE
 Maize Genomics Consortium
 JOURNAL
 Unpublished (2003)

Other GSSs: PUJAX58TD

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TR
 Class: sheared ends.

FEATURES

source
 1. .881
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBTa0626Li6"
 /clone_lib="ZM 0.6 1.0 KB"
 /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN

Query Match 91.6%; Score 17.4; DB 29; Length 881;
 Best Local Similarity 94.7%; Pred. No. 2.8e+04;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGACC 19
 |||||
 Db 457 GTCGACGCCGCGATGACC 475

RESULT 9

CC007495/c
 LOCUS
 DEFINITION
 CC007495 891 bp DNA linear GSS 31-MAR-2003
 PUDHL81TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTa188M18,
 genomic survey sequence.

ACCESSION
 CC007495
 VERSION
 CC007495.1 GI:29386058
 KEYWORDS
 GSS.

SOURCE

Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

TITLE
 Maize Genomics Consortium
 JOURNAL
 Unpublished (2003)

COMMENT
 Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TR
 Class: sheared ends.

FEATURES

source
 1. .891
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBTa188M18"
 /clone_lib="ZM 0.6 1.0 KB"
 /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN

Query Match 91.6%; Score 17.4; DB 28; Length 891;
 Best Local Similarity 94.7%; Pred. No. 2.8e+04;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGACC 19
 |||||
 Db 521 GTCGACGCCGCGATGACC 503

RESULT 10
CB003069
LOCUS
DEFINITION VVB024A10.132996 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB024A10 5', mRNA sequence.
CB003069
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM Vitis vinifera
REFERENCE
AUTHORS Cramer, G.R. and Cushman, J.C.
TITLE An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay
JOURNAL
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 0 row: A column: 10
Seq primer: T3 20mer
High quality sequence stop: 473.
Location/Qualifiers
1..473
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVB024A10"
/tissue_type="leaf"
/dev_stage="juvenile and adult"
/clone_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."

Query Match 89.5%; Score 17; DB 14; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.6e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17
|||||
Db 143 GTCGACGCCGCCGATGA 159

RESULT 11
BJ279225/c
LOCUS
DEFINITION BJ279225 Y. Ogiwara unpublished cDNA library, Wh_r Triticum aestivum cDNA clone whrio09 5', mRNA sequence.
BJ279225
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM Triticum aestivum (bread wheat)
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Pooidae; Triticeae; Triticum.
1 (bases 1 to 571)

AUTHORS TITLE JOURNAL COMMENT

Ogiwara, Y. and Murai, K.
Expressed genes in Triticum aestivum
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1..571
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whrio09"
/tissue_type="root"
/dev_stage="Feekes' scale 1"
/clone_lib="Y. Ogiwara unpublished cDNA library, Wh_r"

FEATURES source

ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 571;
Best Local Similarity 100.0%; Pred. No. 3.7e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGACGCCGCCGATGACC 19
|||||
Db 375 CGACGCCGCCGATGACC 359

RESULT 12 CB003713 LOCUS DEFINITION

CB003713
VVB032C07.134284 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB032C07 5', mRNA sequence.
CB003713
VERSION
KEYWORDS
SOURCE
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
1 (bases 1 to 651)
Cramer, G.R. and Cushman, J.C.
AUTHORS An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay
Unpublished (2002)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 0 row: C column: 07
Seq primer: T3 20mer
High quality sequence stop: 651.
Location/Qualifiers
1..651
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVB032C07"
/tissue_type="leaf"
/dev_stage="juvenile and adult"
/clone_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

CB003713
VVB032C07.134284 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB032C07 5', mRNA sequence.
CB003713
VERSION
KEYWORDS
SOURCE
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
1 (bases 1 to 651)
Cramer, G.R. and Cushman, J.C.
AUTHORS An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay
Unpublished (2002)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 0 row: C column: 07
Seq primer: T3 20mer
High quality sequence stop: 651.
Location/Qualifiers
1..651
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVB032C07"
/tissue_type="leaf"
/dev_stage="juvenile and adult"
/clone_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"

JOURNAL COMMENT

Query Match 89.5%; Score 17; DB 14; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.6e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17
|||||
Db 143 GTCGACGCCGCCGATGA 159

RESULT 11 BJ279225/c LOCUS DEFINITION

BJ279225
BJ279225 Y. Ogiwara unpublished cDNA library, Wh_r Triticum aestivum cDNA clone whrio09 5', mRNA sequence.
BJ279225
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM Triticum aestivum (bread wheat)
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Pooidae; Triticeae; Triticum.
1 (bases 1 to 571)

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

BJ279225
BJ279225 Y. Ogiwara unpublished cDNA library, Wh_r Triticum aestivum cDNA clone whrio09 5', mRNA sequence.
BJ279225
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM Triticum aestivum (bread wheat)
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Pooidae; Triticeae; Triticum.
1 (bases 1 to 571)

/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."

ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 651;
Best Local Similarity 100.0%; Pred. No. 3.7e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATCA 17
|||||
Db 143 GTCGACGCCGCCGATCA 159

RESULT 13

CB003776

LOCUS
DEFINITION VVB033A12.134410 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB033A12 5, mRNA sequence.

ACCESSION CB003776

VERSION CB003776.1 GI:27581081

KEYWORDS EST.

SOURCE Vitis vinifera

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

REFERENCE 1 (bases 1 to 651)

AUTHORS Cramer,G.R. and Cushman,J.C.

TITLE An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay

JOURNAL Unpublished (2002)

COMMENT

Contact: Cushman JC

Department of Biochemistry

University of Nevada

MS200, Reno, NV 89557-0014, USA

Tel: 775-784-1918

Fax: 775-784-1650

Email: jcushman@unr.edu

PCR Primers

FORWARD: T3 20mer

BACKWARD: T7 21mer

Plate: 0 row: A column: 12

Seq primer: T3 20mer

High quality sequence stop: 651.

FEATURES

source

1..651
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"

/clone="VVB033A12"

/tissue_type="leaf"

/dev_stage="juvenile and adult"

/clone_lib="An expressed sequence tag database for abiotic

stressed leaves of Vitis vinifera var. Chardonnay"

/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1:

EcoRI; Site 2: XhoI; Library construction was performed

according to Stratagene's recommended protocol for the

Lambda UniZapXR vector and cDNA synthesis kit."

ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 651;
Best Local Similarity 100.0%; Pred. No. 3.7e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATCA 17
|||||
Db 143 GTCGACGCCGCCGATCA 159

RESULT 14

CB003990

LOCUS

DEFINITION

VVB034E11.134838 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB034E11 5, mRNA sequence.

ACCESSION CB003990

VERSION CB003990.1 GI:27581295

KEYWORDS EST.

SOURCE Vitis vinifera

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

REFERENCE 1 (bases 1 to 653)

AUTHORS Cramer,G.R. and Cushman,J.C.

TITLE An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay

Unpublished (2002)

Contact: Cushman JC

Department of Biochemistry

University of Nevada

MS200, Reno, NV 89557-0014, USA

Tel: 775-784-1918

Fax: 775-784-1650

Email: jcushman@unr.edu

PCR Primers

FORWARD: T3 20mer

BACKWARD: T7 21mer

Plate: 0 row: E column: 11

Seq primer: T3 20mer

High quality sequence stop: 653.

FEATURES

source

1..653
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"

/clone="VVB034E11"

/tissue_type="leaf"

/dev_stage="juvenile and adult"

/clone_lib="An expressed sequence tag database for abiotic

stressed leaves of Vitis vinifera var. Chardonnay"

/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1:

EcoRI; Site 2: XhoI; Library construction was performed

according to Stratagene's recommended protocol for the

Lambda UniZapXR vector and cDNA synthesis kit."

ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 653;
Best Local Similarity 100.0%; Pred. No. 3.7e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATCA 17

Db 143 GTCGACGCCGCCGATCA 159

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp.

FEATURES

source

1. .657
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="wrl109"
/tissue_type="root"
/dev_stage="Feekes' scale 1"
/clone_lib="Y. Ogiwara unpublished cDNA library, Wh_r"

ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 657;
Best Local Similarity 100.0%; Pred. No. 3.7e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGAGCGCCCGCATGACC 19
Db 283 CGAGCGCCCGCATGACC 299

RESULT 16

BE604097/c
LOCUS
DEFINITION
Triticum aestivum cDNA clone WHE1401-1404_E17_E17, mRNA sequence.

ACCESSION
BE604097
VERSION
BE604097.1 GI:9861367
KEYWORDS
EST.

SOURCE
Triticum aestivum (bread wheat)

ORGANISM

Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.

REFERENCE

1 (Bases 1 to 325)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Han, P.S., Hala, C.C.,
Kang, Y., Lazo, G.R., Miller, R., Nguyen H.T., Rausch, C.J.,
Seaton, C.L., Tong, J.C. and Zhang, D.

The structure and function of the expressed portion of the wheat
genomes - Drought stressed leaf cDNA library

JOURNAL

COMMENT
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: andersn@w.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES

source

1. 325
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="TAM W101"
/db_xref="taxon:4565"
/clone="WHE1401-1404_E17_E17"
/tissue_type="Leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_lib="Wheat drought stressed leaf cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were given a gradual
stress down to 65% and 78% RWC at Texas Tech University
(D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA

were prepared, a cDNA library was made, and the cDNA
clones were in vivo excised to give pluescript phagemids
in the TJ Close lab (Choi, Close), at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 325;
Best Local Similarity 94.4%; Pred. No. 5.6e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCATGAC 18
Db 207 GTCGACGCCCGCATGAC 190

RESULT 17

BE409827/c
LOCUS
DEFINITION
1007014F04.2EL_x1 1007 - RescueMu Grid H Zea mays genomic, genomic
survey sequence.

ACCESSION
BE409827
VERSION
BE409827.1 GI:17577672
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 351)
Walbot, V.
Maize genomic sequences found using engineered RescueMu transposon
Unpublished (2001)

JOURNAL

COMMENT
Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Possible ligation site of ends cut by 2 different endonucleases.

Reverse complemented post-ligation sequence from source sequence.

Plate: 1007014 column: 13

Class: transposon-tagged.

FEATURES

source

1. 351
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"

/cultivar="mixed background W23/Al88/B73"

/db_xref="taxon:4577"

/tissue_type="leaf"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="1007 - RescueMu Grid H"

/note="Organ: leaf; Vector: RescueMu (engineered from

pluescript backbone); Site 1: BamHI; Site 2: BglII;

RescueMu is a 4.9 kb, modified maize Mu transposon

designed to allow plasmid rescue from total genomic DNA.

Mu elements insert preferentially into transcription

units. For more information on RescueMu, go to the web

site 'www.zmldb.iastate.edu' and follow the links for

'RescueMu.' Grid H was grown at Berkeley in 2001. DNA

was extracted from leaf punches, double digested using

BamHI and BglII, and ligated to form circular plasmids.

DH10B cells were transformed and then screened on LB

plates with ampicillin."

ORIGIN

Query Match 86.3%; Score 16.4; DB 28; Length 351;
Best Local Similarity 94.4%; Pred. No. 5.6e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query Match 86.3%; Score 16.4; DB 13; Length 414;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19
 |||||
 Db 395 TCAACGCCGCCGATGACC 412

RESULT 21
 CC179114/c
 LOCUS
 DEFINITION SALK_058632.37.95.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_058632.37.95.x, genomic
 survey sequence.

ACCESSION
 CC179114
 VERSION
 CC179114.1 GI:30317665
 KEYWORDS
 SOURCE
 ORGANISM Arabidopsis thaliana (thale cress)

REFERENCE
 1 (bases 1 to 417)
 AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
 Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
 Shinn,P., Zimmerman,J. and Ecker,J.R.
 TITLE A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: eckers@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA.

FEATURES
 source
 Location/Qualifiers
 1..417
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_058632.37.95.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
 Query Match 86.3%; Score 16.4; DB 28; Length 417;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
 |||||
 Db 236 GTCGTCGCCGCCGATGAC 219

RESULT 22
 BF481686/c
 LOCUS
 DEFINITION FMI 21_B06.b1 A003 Floral-Induced Meristem 1 (FMI) Sorghum
 propinquum cDNA, mRNA sequence.

ACCESSION
 BF481686
 VERSION
 BF481686.1 GI:11552507

KEYWORDS
 SOURCE
 ORGANISM Sorghum propinquum
 Sorghum propinquum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE
 1 (bases 1 to 419)
 AUTHORS Cordonnier-Pratt,M.-X., Gingle,A., Sudman,M., Marsala,C. and
 Pratt,L.H.

TITLE An EST database from Sorghum: floral-induced meristems
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@cuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.

Seq primer: JEN REV
 High quality sequence stop: 406
 POLYA=No. Location/Qualifiers
 1..419
 /organism="Sorghum propinquum"
 /mol_type="mRNA"
 /db_xref="taxon:132711"
 /clone_lib="Floral-Induced Meristem 1 (FMI)"
 /note="Organ: Floral-induced meristems; Vector:
 pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
 pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
 EcoRI; mature plants were placed in a growth chamber for
 15 days with 16 hr darkness and 8 hr light (flowering is
 induced by short-day conditions); 16 days after being
 returned to the greenhouse under natural long days during
 late April/early May, meristems were harvested. The
 library was made from poly-A RNA in the cloning vector
 lambda ZAP II. Clones to be sequenced were prepared by
 mass excision."

ORIGIN
 Query Match 86.3%; Score 16.4; DB 10; Length 419;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19
 |||||
 Db 389 TCGACGCCGCCGATGACC 372

RESULT 23
 CG859216
 LOCUS
 DEFINITION ZMMBRC0268G15F ZMMBRC (EcoRI) Zea mays subsp. mays genomic clone
 ZMMBRC0268G15 5', genomic survey sequence.

ACCESSION
 CG859216
 VERSION
 CG859216.1 GI:38431929
 KEYWORDS
 SOURCE
 ORGANISM Zea mays subsp. mays (maize)
 Zea mays subsp. mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
 1 (bases 1 to 425)
 AUTHORS Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
 Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
 TITLE Sequencing of the maize genome at PGIR (2003c)
 JOURNAL Unpublished (2003)
 COMMENT Contact: Bharti,A.K.
 Dr. Joachim Messing's lab
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
 University

190 Frelinghuysen Road, Piscataway, NJ 08854, USA

Tel: 732 445 3801

Fax: 732 445 5735

Email: bharti@aksman.rutgers.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 133.

Location/Qualifiers

1..425
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMWBBC0268G15"
/lab_host="E. coli DH10B"
/clone_lib="ZMWBBC (EcoRI)"
/note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN

Query Match 86.3%; Score 16.4; DB 29; Length 425;
Best Local Similarity 94.4%; Pred. No. 5.7e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCGTGACC 19

|||||

Db 357 TCGACGCCCGCGTGACC 374

RESULT 24

LOCUS

C72864 430 bp mRNA linear EST 04-APR-2002
C72864 Rice panicle at flowering stage Oryza sativa (japonica
cultivar-group) cDNA clone E2385_1A, mRNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sasaki T. and Yamamoto K.
Rice cDNA from panicle at flowering stage
Unpublished (1996)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'.

FEATURES

source

1..430
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E2385_1A"
/dev_stage="flowering stage"
/clone_lib="Rice panicle at flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 430;
Best Local Similarity 94.4%; Pred. No. 5.7e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCGTGACC 19

|||||

Db 218 TCGACGCCCGCGTGACC 235

RESULT 25

LOCUS

DEFINITION

BG465207 440 bp mRNA linear EST 20-MAR-2001
EM1.73.B05.g1.A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor (sorghum)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 440)
Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: PolyTwix

High quality sequence start: 5

High quality sequence stop: 437

POLYA=No.

FEATURES

source

1..440
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A-RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

ORIGIN

Query Match 86.3%; Score 16.4; DB 12; Length 440;
Best Local Similarity 94.4%; Pred. No. 5.7e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCGTGACC 19

|||||

Db 84 TCGACGCCCGCGTGACC 67

RESULT 26

LOCUS

DEFINITION

CA098694 441 bp mRNA linear EST 23-SEP-2003
SCMCCCL6060H11.g CL6 Saccharum officinarum cDNA clone SCMCCCL6060H11
5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Saccharum officinarum
Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.

REFERENCE

AUTHORS

TITLE

JOURNAL

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT

Contact: Arruda P
 Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089
 Email: parruda@unicamp.br
 Clone distribution: clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
<http://www.bcccnet.fcav.unesp.br>
 Plate: 060 row: H column: 11
 Seg primer: T7 Promoter Primer.
 Location/Qualifiers
 1. .441
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCMCL6060H11"
 /lab_host="DH10B"
 /clone_lib="CLJ6"
 /note="Organ: Pool of sugarcane calli submitted to low
 (40C) and high (37 C) temperature stress; Vector: pSport1;
 Site 1: SalI; Site 2: NotI; An unidirectional cDNA library
 generated from [Pool of sugarcane calli submitted to low
 (40C) and high (37 C) temperature stress]. cDNA was
 prepared from polyA+ mRNA using SuperScript Plasmid System
 Kit (Invitrogen). The double-strand cDNAs were
 fractionated in a sepharose CL-2B 40cm-columns and
 fragments sizing between 0.8 and 1.5 Kb were
 directionally cloned into the vector. Details of each
 source of RNA and library construction can be obtained at
<http://sucest.lad.ic.unicamp.br/public>"

FEATURES

source

1. .441
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCMCL6060H11"
 /lab_host="DH10B"
 /clone_lib="CLJ6"
 /note="Organ: Pool of sugarcane calli submitted to low
 (40C) and high (37 C) temperature stress; Vector: pSport1;
 Site 1: SalI; Site 2: NotI; An unidirectional cDNA library
 generated from [Pool of sugarcane calli submitted to low
 (40C) and high (37 C) temperature stress]. cDNA was
 prepared from polyA+ mRNA using SuperScript Plasmid System
 Kit (Invitrogen). The double-strand cDNAs were
 fractionated in a sepharose CL-2B 40cm-columns and
 fragments sizing between 0.8 and 1.5 Kb were
 directionally cloned into the vector. Details of each
 source of RNA and library construction can be obtained at
<http://sucest.lad.ic.unicamp.br/public>"

ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 441;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGAC 18

Db 184 GTCGACGCGCGCGATGAC 167

RESULT 27

C27235/c

LOCUS C27235 Rice callus cDNA Oryza sativa (japonica cultivar-group) cDNA
 DEFINITION clone C51380_2A, mRNA sequence.

ACCESSION C27235

VERSION C27235.1 GI:2311080

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 458)

AUTHORS Yamamoto,K. and Sasaki,T.

TITLE Rice cDNA from callus 1997

JOURNAL Unpublished (1997)

COMMENT Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan

Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@abr.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>

PROJECT = RGP.

Location/Qualifiers

1. .458

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

FEATURES

source

1. .458
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"

/db_xref="taxon:39947"
 /clone="C51380_2A"
 /tissue_type="callus"
 /dev_stage="callus"
 /clone_lib="Rice callus cDNA"

ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 458;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGAC 18

Db 332 GTCGACGCGCGCGATGAC 315

RESULT 28

CG342674/c

LOCUS CG342674

DEFINITION CG342674 468 bp DNA linear GSS 26-AUG-2003
 genomic survey sequence.

ACCESSION CG342674

VERSION CG342674.1 GI:34259940

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 468)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,

Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

Other_GSSs: OGWAJ86TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .468

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMEMA0519004"

/clone_lib="ZM.0.7.1.5_KB"

/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 86.3%; Score 16.4; DB 29; Length 468;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCGCGCGATGACC 19

Db 221 TCGACGCGCGCGATGACC 204

RESULT 29

CF019681/c

LOCUS CF019681

DEFINITION QBN21e01.xg QBN Zea mays cDNA clone QBN21e01, mRNA sequence.

ACCESSION CF019681

VERSION CF019681.1 GI:32914869

KEYWORDS EST.

SOURCE Zea mays

476 bp mRNA linear EST 17-JUL-2003

ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 476)
 Genoplante.
 Genoplante, a major partnership french program in plant genomics
 Unpublished (2003)
 Contact: Genoplante
 Genoplante
 93, rue Henri Rochefort 91025 EVRY CEDEX France
 Tel: 33 1 69 47 54 00
 Fax: 33 1 69 47 54 10
 This sequence has been generated in the framework of the french
 plant genomics programme 'genoplante' (<http://www.genoplante.com>
 and <http://genoplante-info.infobiogen.fr>).

FEATURES

source
 1..476.
 /organism="Zea mays"
 /mol_type="mRNA"
 /cultivar="P2"
 /db_xref="taxon:4577"
 /clone="QBN21e01"
 /tissue_type="pedicel, whole kernel"
 /clone_lib="QBN"

ORIGIN

Query Match 86.3%; Score 16.4; DB 14; Length 476;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19
 |||||
 Db 434 TCGACGCCGCCGATGACC 417

RESULT 30
 BU098652/C
 LOCUS 946129E03.y3 946 - tassell primordium prepared by schmidt lab Zea
 DEFINITION 477 bp mRNA linear EST 29-AUG-2002
 mays cDNA, mRNA sequence.
 ACCESSION BU098652
 VERSION BU098652.1 GI:22546341
 KEYWORDS EST.
 SOURCE Zea mays

ORGANISM

Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 477)

REFERENCE

AUTHORS Walbot V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 JOURNAL University
 COMMENT Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 946129 row: E column: 03.

FEATURES

source
 1..477
 /organism="Zea mays"
 /mol_type="mRNA"
 /cultivar="OH43"
 /db_xref="taxon:4577"
 /tissue_type="tassels"
 /dev_stage="just after the transition from vegetative to
 inflorescence development"
 /lab_host="XLOLR"
 /clone_lib="946 - tassell primordium prepared by Schmidt"

lab"
 /note="Organ: tassels; Vector: HybrizAP; Site 1: EcoRI;
 Site 2: XhoI; George Chuck dissected immature tassels
 between 1mm and 3mm. Sharon Stanfield prepared the cDNA
 library in HybrizAP. Sample insert size range was 350 bp
 to 3 Kb with a 1 Kb average."

ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 477;
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19
 |||||
 Db 436 TCGACGCCGCCGATGACC 419

Search completed: June 20, 2004, 14:15:09
 Job time : 911.873 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 21.0106 Seconds
(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19

Sequence: 1 gtagcgcgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| 1 | 19 | 100.0 | 401 | 3 | US-08-990-823-75 |
| 2 | 19 | 100.0 | 401 | 4 | US-09-477-135A-75 |
| 3 | 19 | 100.0 | 4403765 | 3 | US-09-103-840A-2 |
| 4 | 19 | 100.0 | 4411529 | 3 | US-09-103-840A-1 |
| 5 | 16.4 | 86.3 | 1560 | 4 | US-09-598-747-26 |
| 6 | 16.4 | 86.3 | 4403765 | 3 | US-09-103-840A-2 |
| 7 | 16.4 | 85.3 | 4411529 | 3 | US-09-103-840A-1 |
| 8 | 15.8 | 83.2 | 372 | 4 | US-09-252-991A-14184 |
| 9 | 15.8 | 83.2 | 459 | 4 | US-09-252-991A-4438 |
| 10 | 15.8 | 83.2 | 486 | 4 | US-09-252-991A-1571 |
| 11 | 15.8 | 83.2 | 486 | 4 | US-09-252-991A-10164 |
| 12 | 15.8 | 83.2 | 528 | 4 | US-09-252-991A-4948 |
| 13 | 15.8 | 83.2 | 582 | 4 | US-09-252-991A-1450 |
| 14 | 15.8 | 83.2 | 696 | 4 | US-09-252-991A-4312 |
| 15 | 15.8 | 83.2 | 705 | 4 | US-09-252-991A-5097 |
| 16 | 15.8 | 83.2 | 723 | 4 | US-09-252-991A-5099 |
| 17 | 15.8 | 83.2 | 741 | 4 | US-09-252-991A-11040 |
| 18 | 15.8 | 83.2 | 861 | 4 | US-09-252-991A-13381 |
| 19 | 15.8 | 83.2 | 906 | 4 | US-09-252-991A-9912 |
| 20 | 15.8 | 83.2 | 909 | 4 | US-09-252-991A-1620 |
| 21 | 15.8 | 83.2 | 921 | 4 | US-09-489-039A-1939 |
| 22 | 15.8 | 83.2 | 936 | 4 | US-09-252-991A-4541 |
| 23 | 15.8 | 83.2 | 939 | 4 | US-09-252-991A-13966 |
| 24 | 15.8 | 83.2 | 951 | 4 | US-09-252-991A-16343 |
| 25 | 15.8 | 83.2 | 969 | 4 | US-09-252-991A-15940 |
| 26 | 15.8 | 83.2 | 1002 | 4 | US-09-252-991A-11214 |
| 27 | 15.8 | 83.2 | 1047 | 4 | US-09-252-991A-13479 |

| | | | | | | |
|---|----------------------|-------|------|------|-----|--------------------|
| 4 | US-09-252-991A-1546 | 1068 | 83.2 | 15.8 | 28 | Sequence 1546, Ap |
| 4 | US-09-252-991A-16146 | 1119 | 83.2 | 15.8 | 29 | Sequence 16146, A |
| 4 | US-09-252-991A-10005 | 1902 | 83.2 | 15.8 | 30 | Sequence 10005, A |
| 4 | US-09-252-991A-1449 | 2016 | 83.2 | 15.8 | 31 | Sequence 1449, Ap |
| 4 | US-09-252-991A-1573 | 2022 | 83.2 | 15.8 | 32 | Sequence 1573, Ap |
| 4 | US-09-252-991A-13998 | 2208 | 83.2 | 15.8 | 33 | Sequence 13998, A |
| 4 | US-09-252-991A-14139 | 2439 | 83.2 | 15.8 | 34 | Sequence 14139, A |
| 4 | US-09-252-991A-1619 | 4293 | 83.2 | 15.8 | 35 | Sequence 1619, Ap |
| 4 | US-09-252-991A-1547 | 4359 | 83.2 | 15.8 | 36 | Sequence 1547, Ap |
| 6 | 5183745-1 | 6449 | 83.2 | 15.8 | 37 | Patent No. 5183745 |
| 6 | 5183745-5 | 6441 | 83.2 | 15.8 | 38 | Sequence 1, Appl1 |
| 6 | 5183745-5 | 6443 | 83.2 | 15.8 | 39 | Patent No. 5183745 |
| 4 | US-09-252-991A-1761 | 81.1 | 81.1 | 15.4 | 40 | Sequence 1761, Ap |
| 4 | US-09-489-039A-1291 | 972 | 81.1 | 15.4 | 41 | Sequence 1291, Ap |
| 4 | US-09-489-039A-3159 | 1032 | 81.1 | 15.4 | 42 | Sequence 3159, Ap |
| 4 | US-09-252-991A-2048 | 1191 | 81.1 | 15.4 | 43 | Sequence 2048, Ap |
| 4 | US-09-252-991A-5058 | 1647 | 81.1 | 15.4 | 44 | Sequence 5058, Ap |
| 4 | US-09-186-276B-55 | 1928 | 81.1 | 15.4 | 45 | Sequence 55, Appl |
| 4 | US-08-842-445-55 | 1928 | 81.1 | 15.4 | 46 | Sequence 55, Appl |
| 4 | US-09-186-188B-55 | 1928 | 81.1 | 15.4 | 47 | Sequence 55, Appl |
| 1 | US-08-261-206A-71 | 3306 | 81.1 | 15.4 | 48 | Sequence 71, Appl |
| 4 | US-09-252-991A-1909 | 3546 | 81.1 | 15.4 | 49 | Sequence 1909, Ap |
| 4 | US-09-410-551B-1 | 77536 | 77.9 | 15.8 | 50 | Sequence 1, Appl1 |
| 4 | US-09-313-294A-2856 | 271 | 77.9 | 15.8 | 51 | Sequence 2856, Ap |
| 4 | US-09-313-294A-6425 | 282 | 77.9 | 15.8 | 52 | Sequence 6425, Ap |
| 4 | US-09-252-991A-5579 | 297 | 77.9 | 15.8 | 53 | Sequence 5579, Ap |
| 4 | US-09-313-294A-6043 | 298 | 77.9 | 15.8 | 54 | Sequence 6043, Ap |
| 4 | US-09-252-991A-7007 | 336 | 77.9 | 15.8 | 55 | Sequence 7007, Ap |
| 4 | US-09-252-991A-4191 | 375 | 77.9 | 15.8 | 56 | Sequence 4191, Ap |
| 4 | US-09-489-039A-5141 | 543 | 77.9 | 15.8 | 57 | Sequence 5141, Ap |
| 4 | US-09-252-991A-3318 | 558 | 77.9 | 15.8 | 58 | Sequence 3318, Ap |
| 4 | US-09-252-991A-12270 | 558 | 77.9 | 15.8 | 59 | Sequence 12270, A |
| 4 | US-09-252-991A-2760 | 576 | 77.9 | 15.8 | 60 | Sequence 2760, Ap |
| 4 | US-09-252-991A-4339 | 657 | 77.9 | 15.8 | 61 | Sequence 4339, Ap |
| 4 | US-09-252-991A-1171 | 728 | 77.9 | 15.8 | 62 | Sequence 1171, Ap |
| 4 | US-09-252-991A-1302 | 753 | 77.9 | 15.8 | 63 | Sequence 1302, Ap |
| 4 | US-09-489-039A-2858 | 780 | 77.9 | 15.8 | 64 | Sequence 2858, Ap |
| 3 | US-09-010-809-11 | 801 | 77.9 | 15.8 | 65 | Sequence 11, Appl |
| 4 | US-09-252-991A-15907 | 871 | 77.9 | 15.8 | 66 | Sequence 15907, A |
| 4 | US-09-489-039A-5203 | 822 | 77.9 | 15.8 | 67 | Sequence 5203, Ap |
| 4 | US-09-252-991A-8161 | 918 | 77.9 | 15.8 | 68 | Sequence 8161, Ap |
| 3 | US-09-147-803-21 | 933 | 77.9 | 15.8 | 69 | Sequence 21, Appl |
| 4 | US-09-252-991A-4436 | 939 | 77.9 | 15.8 | 70 | Sequence 4436, Ap |
| 4 | US-09-252-991A-6979 | 963 | 77.9 | 15.8 | 71 | Sequence 6979, Ap |
| 4 | US-09-540-014-27 | 966 | 77.9 | 15.8 | 72 | Sequence 27, Appl |
| 4 | US-09-252-991A-2188 | 966 | 77.9 | 15.8 | 73 | Sequence 2188, Ap |
| 4 | US-09-252-991A-4170 | 999 | 77.9 | 15.8 | 74 | Sequence 4170, Ap |
| 4 | US-09-252-991A-9502 | 999 | 77.9 | 15.8 | 75 | Sequence 9502, Ap |
| 4 | US-09-252-991A-9555 | 1011 | 77.9 | 15.8 | 76 | Sequence 9555, Ap |
| 4 | US-09-252-991A-11274 | 1014 | 77.9 | 15.8 | 77 | Sequence 11274, A |
| 4 | US-09-252-991A-10259 | 1086 | 77.9 | 15.8 | 78 | Sequence 10259, A |
| 4 | US-09-252-991A-10597 | 1101 | 77.9 | 15.8 | 79 | Sequence 10597, A |
| 4 | US-09-252-991A-12132 | 1122 | 77.9 | 15.8 | 80 | Sequence 12132, A |
| 4 | US-09-522-714-13 | 1140 | 77.9 | 15.8 | 81 | Sequence 13, Appl |
| 4 | US-09-252-991A-4515 | 1155 | 77.9 | 15.8 | 82 | Sequence 4515, Ap |
| 2 | US-08-169-948B-13 | 1201 | 77.9 | 15.8 | 83 | Sequence 13, Appl |
| 3 | US-08-488-873-13 | 1201 | 77.9 | 15.8 | 84 | Sequence 13, Appl |
| 3 | US-08-382-452D-13 | 1201 | 77.9 | 15.8 | 85 | Sequence 13, Appl |
| 4 | US-08-507-362A-5 | 1201 | 77.9 | 15.8 | 86 | Sequence 5, Appl1 |
| 4 | US-09-916-494A-13 | 1201 | 77.9 | 15.8 | 87 | Sequence 13, Appl |
| 4 | US-09-252-991A-2489 | 1245 | 77.9 | 15.8 | 88 | Sequence 2489, Ap |
| 4 | US-09-425-578-3 | 1309 | 77.9 | 15.8 | 89 | Sequence 3, Appl1 |
| 4 | US-09-252-991A-9695 | 1332 | 77.9 | 15.8 | 90 | Sequence 9695, Ap |
| 4 | US-09-252-991A-3250 | 1332 | 77.9 | 15.8 | 91 | Sequence 3250, Ap |
| 4 | US-09-489-039A-2930 | 1431 | 77.9 | 15.8 | 92 | Sequence 2930, Ap |
| 4 | US-09-252-991A-2347 | 1431 | 77.9 | 15.8 | 93 | Sequence 2347, Ap |
| 4 | US-09-489-039A-943 | 1470 | 77.9 | 15.8 | 94 | Sequence 943, App |
| 1 | US-08-380-438-1 | 1527 | 77.9 | 15.8 | 95 | Sequence 1, Appl1 |
| 4 | US-09-252-991A-3013 | 1533 | 77.9 | 15.8 | 96 | Sequence 3013, Ap |
| 4 | US-09-252-991A-7033 | 1557 | 77.9 | 15.8 | 97 | Sequence 7033, Ap |
| 4 | US-09-489-039A-2196 | 1575 | 77.9 | 15.8 | 98 | Sequence 2196, Ap |
| 2 | US-08-389-564B-16 | 1588 | 77.9 | 15.8 | 99 | Sequence 16, Appl |
| 3 | US-08-456-047B-16 | 1588 | 77.9 | 15.8 | 100 | Sequence 16, Appl |


```
247 14.2 74.7 594 4 US-09-252-991A-9966 Sequence 9966, Ap
C 248 14.2 74.7 603 4 US-09-489-039A-15 Sequence 15, Appl
C 249 14.2 74.7 606 4 US-09-252-991A-310 Sequence 310, Appl
C 250 14.2 74.7 622 3 US-08-818-112-10 Sequence 10, Appl
C 251 14.2 74.7 622 4 US-08-818-111-10 Sequence 10, Appl
C 252 14.2 74.7 622 4 US-09-056-556-10 Sequence 10, Appl
C 253 14.2 74.7 622 4 US-09-072-956-10 Sequence 10, Appl
C 254 14.2 74.7 622 4 US-09-072-967-10 Sequence 10, Appl
C 255 14.2 74.7 624 4 US-09-252-991A-5716 Sequence 5716, A
C 256 14.2 74.7 627 4 US-09-252-991A-11427 Sequence 11427, A
C 257 14.2 74.7 639 4 US-09-252-991A-2342 Sequence 2342, Ap
C 258 14.2 74.7 645 4 US-09-252-991A-5596 Sequence 5596, Ap
C 259 14.2 74.7 672 4 US-09-252-991A-2494 Sequence 2494, Ap
C 260 14.2 74.7 681 4 US-09-252-991A-10819 Sequence 10819, A
C 261 14.2 74.7 687 4 US-09-252-991A-2644 Sequence 2644, Ap
C 262 14.2 74.7 732 4 US-09-252-991A-2440 Sequence 2440, Ap
C 263 14.2 74.7 747 4 US-09-252-991A-11489 Sequence 11489, A
C 264 14.2 74.7 756 3 US-09-248-335-49 Sequence 49, Appl
C 265 14.2 74.7 777 4 US-09-252-991A-13091 Sequence 13091, A
C 266 14.2 74.7 783 4 US-09-252-991A-6659 Sequence 6659, Ap
C 267 14.2 74.7 795 4 US-09-252-991A-344 Sequence 344, Appl
C 268 14.2 74.7 801 4 US-09-252-991A-15941 Sequence 15941, A
C 269 14.2 74.7 804 4 US-09-252-991A-2187 Sequence 2187, A
C 270 14.2 74.7 819 4 US-09-252-991A-13498 Sequence 13498, A
C 271 14.2 74.7 819 4 US-09-489-039A-6248 Sequence 6248, Ap
C 272 14.2 74.7 822 4 US-09-252-991A-15834 Sequence 15834, A
C 273 14.2 74.7 825 4 US-09-252-991A-5418 Sequence 5418, Ap
C 274 14.2 74.7 834 4 US-09-252-991A-11889 Sequence 11889, A
C 275 14.2 74.7 837 4 US-09-252-991A-8357 Sequence 8357, Ap
C 276 14.2 74.7 870 4 US-09-252-991A-2973 Sequence 2973, Ap
C 277 14.2 74.7 870 4 US-09-489-039A-3614 Sequence 3614, Ap
C 278 14.2 74.7 882 3 US-08-818-112-138 Sequence 138, Appl
C 279 14.2 74.7 882 4 US-08-818-111-133 Sequence 133, Appl
C 280 14.2 74.7 882 4 US-09-056-556-138 Sequence 138, Appl
C 281 14.2 74.7 882 4 US-09-072-956-133 Sequence 133, Appl
C 282 14.2 74.7 882 4 US-09-072-967-138 Sequence 138, Appl
C 283 14.2 74.7 900 4 US-09-252-991A-3544 Sequence 3544, Ap
C 284 14.2 74.7 906 4 US-09-252-991A-1857 Sequence 1857, Ap
C 285 14.2 74.7 909 4 US-09-252-991A-11391 Sequence 11391, A
C 286 14.2 74.7 918 4 US-09-252-991A-9633 Sequence 9633, Ap
C 287 14.2 74.7 930 4 US-09-252-991A-1404 Sequence 1404, Ap
C 288 14.2 74.7 933 4 US-09-252-991A-6169 Sequence 6169, Ap
C 289 14.2 74.7 939 4 US-09-252-991A-13394 Sequence 13394, A
C 290 14.2 74.7 945 4 US-09-252-991A-1918 Sequence 1918, A
C 291 14.2 74.7 948 4 US-09-252-991A-11445 Sequence 11445, A
C 292 14.2 74.7 957 1 US-07-688-299-2 Sequence 2, Appl
C 293 14.2 74.7 957 1 US-07-980-517A-2 Sequence 2, Appl
C 294 14.2 74.7 967 3 US-09-252-991A-6297 Sequence 6297, Ap
C 295 14.2 74.7 967 3 US-09-248-335-51 Sequence 51, Appl
C 296 14.2 74.7 972 4 US-09-252-991A-13601 Sequence 13601, A
C 297 14.2 74.7 993 4 US-09-252-991A-99 Sequence 99, Appl
C 298 14.2 74.7 993 4 US-09-252-991A-6565 Sequence 6565, Ap
C 299 14.2 74.7 1011 4 US-09-252-991A-1771 Sequence 1771, Ap
C 300 14.2 74.7 1020 4 US-09-252-991A-2773 Sequence 2779, Ap
```

ALIGNMENTS

```
RESULT 1
US-09-990-823-75
; Sequence 75, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; EARLIER FILING DATE: 1997-12-15
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254

RESULT 3
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      100.0%; Score 19; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 4219306 GTCGACGCCGCCGATGACC 4219288

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      100.0%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 4227060 GTCGACGCCGCCGATGACC 4227042

RESULT 5
US-09-598-747-26/c
; Sequence 26, Application US/09598747
; Patent No. 6531648
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Michael B.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Casdaska, Pamela Y.
; TITLE OF INVENTION: GRAIN PROCESSING METHOD AND TRANSGENIC PLANTS USEFUL
; TITLE OF INVENTION: THEREIN
; FILE REFERENCE: A-31383P1
; CURRENT APPLICATION NUMBER: US/09/598,747
; CURRENT FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-598-747-26
```

```
Query Match      86.3%; Score 16.4; DB 4; Length 1560;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 906 GTCGACGCCGCCGATGAC 889

RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      86.3%; Score 16.4; DB 3; Length 4403765;
Best Local Similarity 94.4%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 277380 GACGACGCCGCCGATGAC 277397

RESULT 7
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      86.3%; Score 16.4; DB 3; Length 4411529;
Best Local Similarity 94.4%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 277380 GACGACGCCGCCGATGAC 277397
```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1571
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1571

Query Match      83.2%; Score 15.8; DB 4; Length 486;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 311 GTCGACGCCGCCGATGACC 329

RESULT 11
US-09-252-991A-10164/C
; Sequence 10164, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10164
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10164

Query Match      83.2%; Score 15.8; DB 4; Length 486;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 123 GTCGACGCCGCCGATGACC 105

RESULT 12
US-09-252-991A-4948
; Sequence 4948, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4948
; LENGTH: 528

```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14184
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14184

Query Match      83.2%; Score 15.8; DB 4; Length 372;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 33 GTCGACGCCGCCGATGACC 51

RESULT 9
US-09-252-991A-4438/C
; Sequence 4438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4438
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4438

Query Match      83.2%; Score 15.8; DB 4; Length 459;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 148 GTTCCGCCGCCGATGACC 130

RESULT 10
US-09-252-991A-1571
; Sequence 1571, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```

```
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4948

Query Match      83.2%; Score 15.8; DB 4; Length 528;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 415 GTCGCGCGCGATGACC 433

RESULT 13
US-09-252-991A-1450/c
; Sequence 1450, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1450
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1450

Query Match      83.2%; Score 15.8; DB 4; Length 582;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 74 GTCGACGCGCGATGACC 56

RESULT 14
US-09-252-991A-4312
; Sequence 4312, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4312
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4312

Query Match      83.2%; Score 15.8; DB 4; Length 696;
Best Local Similarity 89.5%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 559 GTTGGCGCGCGATGACC 577

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4948

Query Match      83.2%; Score 15.8; DB 4; Length 528;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 415 GTCGCGCGCGATGACC 433

RESULT 13
US-09-252-991A-1450/c
; Sequence 1450, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1450
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1450

Query Match      83.2%; Score 15.8; DB 4; Length 582;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 74 GTCGACGCGCGATGACC 56

RESULT 14
US-09-252-991A-4312
; Sequence 4312, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4312
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4312

Query Match      83.2%; Score 15.8; DB 4; Length 696;
Best Local Similarity 89.5%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 559 GTTGGCGCGCGATGACC 577

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4948

Query Match      83.2%; Score 15.8; DB 4; Length 723;
Best Local Similarity 89.5%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 384 GTCGCGCGCGATGACC 366

RESULT 17
US-09-252-991A-11040
; Sequence 11040, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
```

US-09-252-991A-9912

Query Match 83.2%; Score 15.8; DB 4; Length 906;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 715 GTCGACGCCGCCGATGACC 733

RESULT 20

US-09-252-991A-1620/c
; Sequence 1620, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1620
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1620

Query Match 83.2%; Score 15.8; DB 4; Length 909;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 497 GTCGACGCCGCCGATGACC 479

RESULT 21

US-09-489-039A-1939/c
; Sequence 1939, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bretton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1939
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1939

Query Match 83.2%; Score 15.8; DB 4; Length 921;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 465 GTCGACATCGCCGATGACC 447

RESULT 22

US-09-252-991A-4541/c

US-09-252-991A-11040

Query Match 83.2%; Score 15.8; DB 4; Length 741;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 178 GTCGACGCCGCCGATGACC 196

RESULT 18

US-09-252-991A-13381/c
; Sequence 13381, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13381
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13381

Query Match 83.2%; Score 15.8; DB 4; Length 861;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 184 GTCGACGCCGCCGATGATC 166

RESULT 19

US-09-252-991A-9912
; Sequence 9912, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9912
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

; Sequence 4541, Application US/09252991A
; Patent No. 6551795

Query Match 83.2%; Score 15.8; DB 4; Length 936;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 23

US-09-252-991A-13966
; Sequence 13966, Application US/09252991A

Query Match 83.2%; Score 15.8; DB 4; Length 939;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels

RESULT 24

US-09-252-991A-16343
; Sequence 16343, Application US/09252991A

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142

Query Match 83.2%; Score 15.8; DB 4; Length 951;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 25

RESULT 25
US-09-252-991A-15940/c
: Sequence 15940. Application US/09252991A

Query Match 83.2%; Score 15.8; DB 4; Length 969;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 26

RESOLUTION 28
US-09-252-991A-11214/c
: Sequence 11214. Application US/09252991A

| | | | | |
|-----------------------|--------|--------------------|-------|--------------|
| Query Match | 83.2%; | Score 15.8; | DB 4; | Length 1002; |
| Best Local Similarity | 89.5%; | Pred. No. 1.9e+02; | | |

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
|||||
Db 705 GTCGACGCCGCCGATGACC 587

RESULT 27

US-09-252-991A-13479/c

; Sequence 13479, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13479

; LENGTH: 1047

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13479

Query Match 83.2%; Score 15.8; DB 4; Length 1047;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
|||||
Db 441 GTCGACGCCGCCGATGATC 423

RESULT 28

US-09-252-991A-1546

; Sequence 1546, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 1546

; LENGTH: 1068

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-1546

Query Match 83.2%; Score 15.8; DB 4; Length 1068;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
|||||
Db 851 GTCGACGCCGCCGATGACC 869

RESULT 29

US-09-252-991A-16146/c

; Sequence 16146, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16146
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16146

Query Match 83.2%; Score 15.8; DB 4; Length 1119;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
|||||
Db 181 GTCGACGCCGCCGATGATC 163

RESULT 30

US-09-252-991A-10005

; Sequence 10005, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 10005

; LENGTH: 1902

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10005

Query Match 83.2%; Score 15.8; DB 4; Length 1902;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
|||||
Db 1836 GTCGACGCCGCCGATGACC 1854

Search completed: June 20, 2004, 11:46:54

Job time : 35.1356 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 107.466 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19

Sequence: 1 gtcgacgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | ID | Description |
|------------|-------|-------|---------|--------|----------------------|--------------------|
| 1 | 19 | 100.0 | 401 | 9 | US-09-996-634-75 | Sequence 75, Appl |
| 2 | 19 | 100.0 | 401 | 10 | US-09-997-182-75 | Sequence 75, Appl |
| 3 | 19 | 100.0 | 401 | 10 | US-09-997-181-75 | Sequence 75, Appl |
| 4 | 17.4 | 91.6 | 1062 | 15 | US-10-156-761-626 | Sequence 626, Appl |
| 5 | 17.4 | 91.6 | 4149 | 15 | US-10-121-988-84 | Sequence 84, Appl |
| 6 | 17.4 | 91.6 | 4149 | 15 | US-10-200-582-84 | Sequence 84, Appl |
| 7 | 17.4 | 91.6 | 4149 | 15 | US-10-237-551-84 | Sequence 84, Appl |
| 8 | 17.4 | 91.6 | 9025608 | 15 | US-10-156-761-1 | Sequence 1, Appl |
| 9 | 17 | 89.5 | 25 | 15 | US-10-099-263B-37960 | Sequence 37960, A |
| 10 | 16.4 | 86.3 | 246 | 13 | US-10-425-114-22954 | Sequence 22954, A |
| 11 | 16.4 | 86.3 | 270 | 9 | US-09-294-093B-1992 | Sequence 1992, Ap |
| 12 | 16.4 | 86.3 | 429 | 16 | US-10-260-238-5854 | Sequence 5854, Ap |
| 13 | 16.4 | 86.3 | 453 | 15 | US-10-156-761-1762 | Sequence 1762, Ap |
| 14 | 16.4 | 86.3 | 646 | 13 | US-10-425-114-15573 | Sequence 15573, A |

| | | | | | |
|------|------|---------|----|----------------------|--------------------|
| 16.4 | 86.3 | 702 | 13 | US-10-424-593-23643 | Sequence 23643, A |
| 16.4 | 86.3 | 836 | 13 | US-10-425-114-2101 | Sequence 2101, Ap |
| 16.4 | 86.3 | 869 | 13 | US-10-425-114-30130 | Sequence 30330, A |
| 16.4 | 86.3 | 880 | 13 | US-10-425-114-16256 | Sequence 16256, A |
| 16.4 | 86.3 | 882 | 13 | US-10-425-114-1435 | Sequence 1435, Ap |
| 16.4 | 86.3 | 882 | 13 | US-10-425-114-12308 | Sequence 12308, A |
| 16.4 | 86.3 | 889 | 13 | US-10-425-114-17287 | Sequence 17287, A |
| 16.4 | 86.3 | 893 | 13 | US-10-425-114-7159 | Sequence 7159, Ap |
| 16.4 | 86.3 | 910 | 13 | US-10-424-593-120250 | Sequence 120250, |
| 16.4 | 86.3 | 917 | 13 | US-10-425-114-3860 | Sequence 3860, Ap |
| 16.4 | 86.3 | 958 | 13 | US-10-425-114-5238 | Sequence 5238, Ap |
| 16.4 | 86.3 | 1077 | 16 | US-10-260-238-313 | Sequence 313, Appl |
| 16.4 | 86.3 | 1232 | 13 | US-10-425-114-5104 | Sequence 5104, Ap |
| 16.4 | 86.3 | 1263 | 16 | US-10-369-493-31945 | Sequence 31945, A |
| 16.4 | 86.3 | 1309 | 13 | US-10-425-114-1473 | Sequence 1473, Ap |
| 16.4 | 86.3 | 1470 | 13 | US-10-425-114-2580 | Sequence 2580, Ap |
| 16.4 | 86.3 | 1558 | 13 | US-10-425-114-32297 | Sequence 32297, A |
| 16.4 | 86.3 | 1560 | 15 | US-10-306-292-25 | Sequence 26, Appl |
| 16.4 | 86.3 | 1686 | 13 | US-10-425-114-3895 | Sequence 30895, A |
| 16.4 | 86.3 | 1763 | 13 | US-10-425-114-32140 | Sequence 32140, A |
| 16.4 | 86.3 | 1935 | 16 | US-10-369-493-32293 | Sequence 32293, A |
| 16.4 | 86.3 | 1948 | 13 | US-10-282-122A-15101 | Sequence 15101, A |
| 16.4 | 86.3 | 2000 | 9 | US-09-938-842A-4146 | Sequence 4146, Ap |
| 16.4 | 86.3 | 2000 | 11 | US-10-425-114-32995 | Sequence 32995, A |
| 16.4 | 86.3 | 2357 | 13 | US-09-923-876-1729 | Sequence 1729, Ap |
| 16 | 84.2 | 234 | 9 | US-09-923-876-1729 | Sequence 812, Appl |
| 16 | 84.2 | 270 | 15 | US-10-156-761-812 | Sequence 13852, A |
| 16 | 84.2 | 821 | 13 | US-10-425-114-13852 | Sequence 35, Appl |
| 16 | 84.2 | 942 | 13 | US-10-440-503-35 | Sequence 35, Appl |
| 16 | 84.2 | 942 | 15 | US-10-461-925-35 | Sequence 35, Appl |
| 16 | 84.2 | 942 | 15 | US-10-146-772-35 | Sequence 35, Appl |
| 16 | 84.2 | 942 | 16 | US-10-241-523-35 | Sequence 35, Appl |
| 16 | 84.2 | 1029 | 15 | US-10-440-523-35 | Sequence 35, Appl |
| 16 | 84.2 | 1041 | 15 | US-10-156-761-1970 | Sequence 1970, Ap |
| 16 | 84.2 | 1053 | 16 | US-10-156-761-7249 | Sequence 7249, Ap |
| 16 | 84.2 | 1381 | 16 | US-10-369-493-27555 | Sequence 27555, A |
| 16 | 84.2 | 1821 | 10 | US-10-369-493-45397 | Sequence 45397, A |
| 16 | 84.2 | 3732 | 16 | US-09-272-975-4 | Sequence 4, Appl |
| 16 | 84.2 | 9025608 | 15 | US-10-369-493-44237 | Sequence 44237, A |
| 15.8 | 83.2 | 345 | 15 | US-10-156-761-388 | Sequence 1, Appl |
| 15.8 | 83.2 | 477 | 13 | US-10-282-122A-30409 | Sequence 388, Appl |
| 15.8 | 83.2 | 600 | 15 | US-10-156-761-6597 | Sequence 30409, A |
| 15.8 | 83.2 | 702 | 15 | US-10-156-761-2104 | Sequence 6597, Ap |
| 15.8 | 83.2 | 738 | 13 | US-10-282-122A-30336 | Sequence 2104, Ap |
| 15.8 | 83.2 | 742 | 16 | US-10-282-122A-30336 | Sequence 30336, A |
| 15.8 | 83.2 | 744 | 13 | US-10-260-238-4028 | Sequence 4028, Ap |
| 15.8 | 83.2 | 762 | 15 | US-10-424-593-40764 | Sequence 40764, A |
| 15.8 | 83.2 | 828 | 9 | US-10-156-761-2385 | Sequence 2385, Ap |
| 15.8 | 83.2 | 828 | 13 | US-09-815-242-7971 | Sequence 7971, Ap |
| 15.8 | 83.2 | 894 | 15 | US-10-282-122A-30566 | Sequence 30566, A |
| 15.8 | 83.2 | 954 | 15 | US-10-156-761-1180 | Sequence 1180, Ap |
| 15.8 | 83.2 | 993 | 13 | US-10-156-761-1018 | Sequence 1018, Ap |
| 15.8 | 83.2 | 1005 | 16 | US-10-282-122A-32095 | Sequence 32095, A |
| 15.8 | 83.2 | 1017 | 15 | US-10-369-493-24202 | Sequence 24202, A |
| 15.8 | 83.2 | 1041 | 13 | US-10-156-761-5970 | Sequence 5970, Ap |
| 15.8 | 83.2 | 1053 | 16 | US-10-282-122A-17694 | Sequence 17694, A |
| 15.8 | 83.2 | 1064 | 13 | US-10-369-493-34377 | Sequence 34377, A |
| 15.8 | 83.2 | 1064 | 17 | US-10-250-721-7 | Sequence 7, Appl |
| 15.8 | 83.2 | 1068 | 13 | US-10-250-824-7 | Sequence 7, Appl |
| 15.8 | 83.2 | 1107 | 16 | US-10-107-431-134 | Sequence 134, Appl |
| 15.8 | 83.2 | 1155 | 13 | US-10-369-493-42249 | Sequence 42249, A |
| 15.8 | 83.2 | 1161 | 15 | US-10-425-114-22383 | Sequence 22383, A |
| 15.8 | 83.2 | 1182 | 13 | US-10-156-761-7033 | Sequence 7033, Ap |
| 15.8 | 83.2 | 1294 | 15 | US-10-156-761-6347 | Sequence 6347, Ap |
| 15.8 | 83.2 | 1350 | 15 | US-10-425-114-22890 | Sequence 22890, A |
| 15.8 | 83.2 | 1366 | 16 | US-10-156-761-3627 | Sequence 3627, Ap |
| 15.8 | 83.2 | 1386 | 16 | US-10-369-493-32374 | Sequence 32374, A |
| 15.8 | 83.2 | 1416 | 16 | US-10-369-493-35492 | Sequence 35492, A |
| 15.8 | 83.2 | 1425 | 16 | US-10-369-493-45481 | Sequence 45481, A |
| 15.8 | 83.2 | 1440 | 16 | US-10-369-493-44445 | Sequence 44445, A |
| 15.8 | 83.2 | 1467 | 9 | US-10-369-493-40480 | Sequence 40480, A |
| 15.8 | 83.2 | | | US-09-738-626-305 | Sequence 305, Appl |

US-09-996-634-75
; Sequence 75, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR FILING DATE: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(401)
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.
US-09-996-634-75

Query Match 100.0%; Score 19; DB 9; Length 401;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCCGCATGACC 19
Db 12 GTCGACGCCCGCATGACC 30

RESULT 2
US-09-997-182-75
; Sequence 75, Application US/09997182
; Publication No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; CURRENT FILING DATE: 2001-11-28
; PRIOR FILING DATE: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(401)
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.
US-09-997-182-75

Query Match 100.0%; Score 19; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 9, Appli
Sequence 29773, A
Sequence 14772, A
Sequence 1401, A
Sequence 14374, A
Sequence 28827, A
Sequence 29519, A
Sequence 339, Appli
Sequence 139738, A
Sequence 12112, A
Sequence 14241, A
Sequence 34345, A
Sequence 23407, A
Sequence 145, App
Sequence 12522, A
Sequence 183436, A
Sequence 28803, A
Sequence 57549, A
Sequence 201, App
Sequence 6930, App
Sequence 1424, App
Sequence 1267, App
Sequence 186, App
Sequence 1995, App
Sequence 6076, App
Sequence 39580, App
Sequence 39593, App
Sequence 31843, App
Sequence 29018, App
Sequence 7843, App
Sequence 28785, App
Sequence 30338, App
Sequence 39209, App
Sequence 42344, App
Sequence 2612, App
Sequence 5022, App
Sequence 30532, App
Sequence 118913, App
Sequence 57527, App
Sequence 31124, App
Sequence 16061, App
Sequence 47111, App
Sequence 23640, App
Sequence 227, App
Sequence 27685, App
Sequence 26536, App
Sequence 1426, App
Sequence 1269, App
Sequence 28738, App
Sequence 23637, App
Sequence 23841, App
Sequence 23192, App
Sequence 5117, App
Sequence 47104, App
Sequence 30966, App
Sequence 27621, App
Sequence 11189, App
Sequence 329, App
Sequence 32531, App
Sequence 31794, App
Sequence 28207, App
Sequence 14278, App
Sequence 5167, App
Sequence 1027, App
Sequence 1609, App
Sequence 21, Appli

US-10-125-258-9
US-10-282-122A-29773
US-10-425-114-14472
US-10-282-122A-11401
US-10-282-122A-14374
US-10-282-122A-28827
US-10-282-122A-29519
US-09-734-569-39
US-10-424-599-139738
US-10-425-114-12112
US-10-425-114-14241
US-10-424-599-34345
US-10-425-114-23407
US-10-425-114-145
US-10-425-114-12522
US-10-027-632-183436
US-10-027-632-28803
US-10-424-599-57549
US-10-389-647-201
US-10-156-761-6930
US-10-412-699B-1424
US-10-374-780A-1267
US-10-260-238-186
US-10-156-761-1995
US-10-156-761-6076
US-10-369-493-39580
US-10-369-493-39593
US-10-282-122A-31843
US-10-425-114-29018
US-09-815-242-7843
US-10-282-122A-28785
US-10-282-122A-30338
US-10-369-493-39209
US-10-156-761-42344
US-10-156-761-2612
US-10-425-114-5022
US-10-424-599-30532
US-10-424-599-118913
US-10-424-599-57527
US-10-425-114-31124
US-10-425-114-16061
US-10-369-493-47111
US-10-425-114-23640
US-09-738-626-227
US-10-425-114-27685
US-10-282-122A-26536
US-10-412-699B-1426
US-10-374-780A-1269
US-10-282-122A-28738
US-10-425-114-23637
US-10-425-114-23841
US-10-425-114-23192
US-10-156-761-5117
US-10-369-493-47104
US-10-369-493-30966
US-10-369-493-27621
US-10-282-122A-11189
US-09-974-300-329
US-10-424-599-32531
US-10-425-114-31794
US-10-369-493-28207
US-10-425-114-14278
US-10-156-761-5167
US-10-425-114-1027
US-10-425-114-1609
US-09-882-631-21

444 77.9 15
447 77.9 13
464 77.9 13
468 77.9 13
468 77.9 13
468 77.9 13
471 77.9 13
480 77.9 13
485 77.9 13
504 77.9 13
512 77.9 13
514 77.9 13
530 77.9 13
539 77.9 13
540 77.9 13
551 77.9 13
551 77.9 13
570 77.9 13
589 77.9 13
606 77.9 13
624 77.9 13
633 77.9 13
633 77.9 13
646 77.9 13
654 77.9 13
675 77.9 13
732 77.9 13
732 77.9 13
735 77.9 13
735 77.9 13
737 77.9 13
741 77.9 13
741 77.9 13
741 77.9 13
747 77.9 13
750 77.9 13
756 77.9 13
759 77.9 13
762 77.9 13
771 77.9 13
773 77.9 13
779 77.9 13
781 77.9 13
783 77.9 13
786 77.9 13
789 77.9 13
806 77.9 13
819 77.9 13
820 77.9 13
820 77.9 13
822 77.9 13
828 77.9 13
830 77.9 13
831 77.9 13
843 77.9 13
852 77.9 13
855 77.9 13
872 77.9 13
875 77.9 13
879 77.9 13
885 77.9 13
886 77.9 13
888 77.9 13
891 77.9 13
893 77.9 13
900 77.9 13
908 77.9 13
927 77.9 13
933 77.9 13

ALIGNMENTS

RESULT 1

```

QY      1  GTCGACGCCCGCGATGACC 19
Db      12 GTCGACGCCCGCGATGACC 30

RESULT 3
US-09-997-181-75
; Sequence 75, Application US/09997181
; Publication No. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Naro, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61257
; CURRENT FILING DATE: 2001-11-28
; PRIOR FILING DATE: 2000-01-03
; PRIOR FILING DATE: 1997-12-15
; PRIOR FILING DATE: 1996-06-14
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; NAME/KEY: misc feature
; LOCATION: (1)..(401)
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.
US-09-997-181-75

```

```

Query Match      100.0%; Score 19; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  GTCGACGCCCGCGATGACC 19
Db      12 GTCGACGCCCGCGATGACC 30

```

```

RESULT 4
US-10-156-761-626/c
; Sequence 626, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2001-05-30
; PRIOR FILING DATE: 2001-05-30
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 626
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; NAME/KEY: CDS

```

```

; LOCATION: (1)..(1062)
US-10-156-761-626
Query Match      91.6%; Score 17.4; DB 15; Length 1062;
Best Local Similarity 94.7%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GTCGACGCCCGCGATGACC 19
Db      36 GTCGACGCCCGCGATGACC 18

```

```

RESULT 5
US-10-121-988-84
; Sequence 84, Application US/10121988
; Publication No. US20030068327A1
; GENERAL INFORMATION:
; APPLICANT: Hosken, Nancy Ann
; APPLICANT: McGowan, Patrick
; APPLICANT: Sleath, Paul R.
; APPLICANT: Mossman, Sally P.
; APPLICANT: Evans, Lawrence S.
; APPLICANT: Swanson, Ryan M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C1
; CURRENT APPLICATION NUMBER: US/10/121,988
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 183
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-121-988-84

```

```

Query Match      91.6%; Score 17.4; DB 15; Length 4149;
Best Local Similarity 94.7%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GTCGACGCCCGCGATGACC 19
Db      2395 GTCGACGCCCGCGATGACC 2413

```

```

RESULT 6
US-10-200-562-84
; Sequence 84, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowan, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-200-562-84

```

```

Query Match      91.6%; Score 17.4; DB 15; Length 4149;
Best Local Similarity 94.7%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GTCGACGCCCGCGATGACC 19
Db      2395 GTCGACGCCCGCGATGACC 2413

```

```
US-10-098-263B-37960
; Sequence 37960, Application US/10098263B
; Publication No. US20030104410A1
; GENERAL INFORMATION:
; APPLICANT: Mittman, Michael
; TITLE OF INVENTION: Human Microarray
; FILE REFERENCE: 3118.1
; CURRENT APPLICATION NUMBER: US/10/098,263B
; PRIOR FILING DATE: 2003-01-08
; PRIOR APPLICATION NUMBER: 60/276,759
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 131066
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 37960
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
; US-10-098-263B-37960

Query Match      89.5%; Score 17; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATCA 17
Db 9 GTCGACGCCGCCGATCA 25

RESULT 10
US-10-425-114-22954/c
; Sequence 22954, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22954
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-081-C10_FLI
; US-10-425-114-22954

Query Match      86.3%; Score 16.4; DB 13; Length 246;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19
Db 105 TCGACGCCGCGTGTGATGACC 88

RESULT 11
US-09-294-093B-1992/c
; Sequence 1992, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
```

```
US-10-237-551-84
; Sequence 84, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-237-551-84

Query Match      91.6%; Score 17.4; DB 15; Length 4149;
Best Local Similarity 94.7%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
Db 2395 GTCGACGCCGCCGATGACC 2413

RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Query Match      91.6%; Score 17.4; DB 15; Length 9025608;
Best Local Similarity 94.7%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
Db 788283 GTCGACGCCGCCGATGACC 788265

RESULT 9
```

; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1992
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700345337H1
; NAME/KEY: unsure
; LOCATION: 68
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-1992

Query Match 86.3%; Score 16.4; DB 9; Length 270;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
|||
Db 87 GTCGACGCCGCCGATGAC 70

RESULT 12
US-10-260-238-5854/c
; Sequence 5854, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 5854
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Zea mays
US-10-260-238-5854

Query Match 86.3%; Score 16.4; DB 16; Length 429;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19
|||
Db 298 TCGACGCCGCCGATGACC 281

RESULT 13
US-10-156-761-1762/c
; Sequence 1762, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1762
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(453)
US-10-156-761-1762

Query Match 86.3%; Score 16.4; DB 15; Length 453;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
|||
Db 372 GTCGACGCCGCCGATGAC 355

RESULT 14
US-10-425-114-15573/c
; Sequence 15573, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15573
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-027-C5_FLI
US-10-425-114-15573

Query Match 86.3%; Score 16.4; DB 13; Length 646;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19
|||
Db 296 TCGACGCCGCCGATGACC 279

RESULT 15
US-10-424-599-23643/c
; Sequence 23643, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 30330
LENGTH: 869
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: UC-ZMFLB73010B10_FLI
US-10-425-114-30330
Query Match 86.3%; Score 16.4; DB 13; Length 869;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCGACGCCGCCGATGACC 19
Db 364 TCGACGCCGCCGATGACC 347
RESULT 18
US-10-425-114-16256/c
Sequence 16256, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 16256
LENGTH: 880
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3062-022-E9_FLI
US-10-425-114-16256
Query Match 86.3%; Score 16.4; DB 13; Length 880;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCGACGCCGCCGATGACC 19
Db 422 TCGACGCCGCCGATGACC 405
RESULT 19
US-10-425-114-1435/c
Sequence 1435, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 23643
LENGTH: 702
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)-(702)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_121351C.1
US-10-424-599-23643
Query Match 86.3%; Score 16.4; DB 13; Length 702;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCGACGCCGCCGATGACC 19
Db 176 TCGACGCCGCCGATGACC 159
RESULT 16
US-10-425-114-2101/c
Sequence 2101, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 2101
LENGTH: 836
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700203248_FLI
US-10-425-114-2101
Query Match 86.3%; Score 16.4; DB 13; Length 836;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TCGACGCCGCCGATGACC 19
Db 343 TCGACGCCGCCGATGACC 326
RESULT 17
US-10-425-114-30330/c
Sequence 30330, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.

```
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1435
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700152962_FLI
US-10-425-114-1435

Query Match      86.3%; Score 16.4; DB 13; Length 882;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      388 TCGACGCCGCCGATGACC 371

RESULT 20
US-10-425-114-12308/c
; Sequence 12308, Application US/10/425,114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 12308
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701166706_FLI
US-10-425-114-12308

Query Match      86.3%; Score 16.4; DB 13; Length 882;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      336 TCGACGCCGCCGATGACC 319

RESULT 21
US-10-425-114-17287/c
; Sequence 17287, Application US/10/425,114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 17287
```

```
; LENGTH: 889
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-003-H4_FLI
US-10-425-114-17287

Query Match      86.3%; Score 16.4; DB 13; Length 889;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      396 TCGACGCCGCCGATGACC 379

RESULT 22
US-10-425-114-7159/c
; Sequence 7159, Application US/10/425,114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 7159
; LENGTH: 893
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700620440_FLI
US-10-425-114-7159

Query Match      86.3%; Score 16.4; DB 13; Length 893;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      395 TCGACGCCGCCGATGACC 378

RESULT 23
US-10-424-599-120250
; Sequence 120250, Application US/10/424,599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 120250
; LENGTH: 910
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_79594C.1
US-10-424-599-120250
```


Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCATGACC 18
|||||
Db 209-GTCGACGCCCGCAAGAC 192

RESULT 28

US-10-369-493-31945/c
; Sequence 31945, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianteng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31945
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Thermobifida fusca
US-10-369-493-31945

Query Match 86.3%; Score 16.4; DB 16; Length 1263;

Best Local Similarity 94.4%; Pred. No. 1.3e-02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCATGACC 19
|||||
Db 236 TCGATGCCCGCGCATGACC 219

RESULT 29

US-10-425-114-1473/c
; Sequence 1473, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1473
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700154547_FLI
US-10-425-114-1473

Query Match 86.3%; Score 16.4; DB 13; Length 1309;

Best Local Similarity 94.4%; Pred. No. 1.3e-02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCATGACC 19
|||||
Db 913 TCGACGCCCGCGTGACC 896

RESULT 30

US-10-425-114-2580/c
; Sequence 2580, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 2580
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700217894_FLI
US-10-425-114-2580

Query Match 86.3%; Score 16.4; DB 13; Length 1470;

Best Local Similarity 94.4%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCATGACC 19
|||||
Db 997 TCGACGCCCGCGTGACC 980

Search completed: June 20, 2004, 17:46:28

Job time : 122.466 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 515.915 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19
Sequence: 1 gtcagccgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_btg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_in:*
- 18: em_mu:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pi:*
- 25: em_ro:*
- 26: em_st:*
- 27: em_un:*
- 28: em_vt:*
- 29: em_vt:*
- 30: em_vt:*
- 31: em_vt:*
- 32: em_vt:*
- 33: em_vt:*
- 34: em_vt:*
- 35: em_vt:*
- 36: em_vt:*
- 37: em_vt:*
- 38: em_vt:*
- 39: em_vt:*
- 40: em_vt:*
- 41: em_vt:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|-------------|
| 1 | 19 | 100.0 | 401 | 6 | AR149059 | Sequence |
| 2 | 19 | 100.0 | 401 | 6 | AR340431 | Sequence |
| C 3 | 19 | 100.0 | 18121 | 1 | AE007182 | Sequence |
| C 4 | 19 | 100.0 | 37164 | 1 | MSGY126 | Sequence |
| C 5 | 19 | 100.0 | 244800 | 1 | EX842584 | Sequence |
| C 6 | 19 | 100.0 | 278492 | 1 | EX248347 | Sequence |
| 7 | 17.4 | 91.6 | 2534 | 8 | AK072452 | Oryza sat |
| C 8 | 17.4 | 91.6 | 5966 | 1 | SAWY18862 | Sequence |
| 9 | 17.4 | 91.6 | 60687 | 8 | EX842638 | Sequence |
| C 10 | 17.4 | 91.6 | 139614 | 8 | AC068950 | Oryza sat |
| 11 | 17.4 | 91.6 | 154288 | 2 | AP005494 | Oryza sat |
| 12 | 17.4 | 91.6 | 230138 | 14 | AF232689 | Sequence |
| C 13 | 17.4 | 91.6 | 299175 | 1 | AF003023 | Sequence |
| 14 | 17.4 | 91.6 | 299850 | 1 | AP005949 | Sequence |
| 15 | 17.4 | 91.6 | 301666 | 8 | AE017108 | Oryza sat |
| 16 | 17.4 | 91.6 | 343473 | 1 | EX640451 | Bordetell |
| 17 | 17.4 | 91.6 | 346259 | 1 | EX640435 | Bordetell |
| 18 | 17.4 | 91.6 | 348171 | 1 | EX640412 | Bordetell |
| C 19 | 17 | 89.5 | 6736 | 14 | ADP279815 | Sequence |
| 20 | 17 | 89.5 | 11252 | 1 | AE005103 | Halobacte |
| C 21 | 17 | 89.5 | 113170 | 8 | AP003804 | Oryza sat |
| C 22 | 17 | 89.5 | 347950 | 1 | AP003013 | Mesorhizo |
| 23 | 16.4 | 86.3 | 227 | 8 | AY023290 | Oryza sat |
| C 24 | 16.4 | 86.3 | 810 | 1 | AFAAZU | Sequence |
| C 25 | 16.4 | 86.3 | 963 | 6 | AX653785 | Sequence |
| C 26 | 16.4 | 86.3 | 1077 | 6 | AX660080 | Sequence |
| C 27 | 16.4 | 86.3 | 1339 | 5 | AB025282 | Callus ga |
| C 28 | 16.4 | 86.3 | 1389 | 8 | AK104317 | Oryza sat |
| C 29 | 16.4 | 86.3 | 1502 | 8 | AK071251 | Oryza sat |
| C 30 | 16.4 | 86.3 | 1560 | 6 | AX366999 | Sequence |
| C 31 | 16.4 | 86.3 | 1796 | 8 | AK065989 | Oryza sat |
| 32 | 16.4 | 86.3 | 2000 | 6 | AX509451 | Sequence |
| 33 | 16.4 | 86.3 | 2105 | 8 | AK071946 | Oryza sat |
| 34 | 16.4 | 86.3 | 2217 | 8 | AK072413 | Oryza sat |
| 35 | 16.4 | 86.3 | 2431 | 1 | AF099189 | Caulobact |
| 36 | 16.4 | 86.3 | 3249 | 8 | AK110103 | Oryza sat |
| C 37 | 16.4 | 86.3 | 5076 | 5 | AB025284 | Gallus ga |
| C 38 | 16.4 | 86.3 | 10352 | 1 | AF546152 | Streptomy |
| C 39 | 16.4 | 86.3 | 10784 | 1 | AE014719 | Bifidobac |
| 40 | 16.4 | 86.3 | 10927 | 1 | AE002535 | Neisseria |
| 41 | 16.4 | 86.3 | 11122 | 1 | AE006056 | Pasteurel |
| 42 | 16.4 | 86.3 | 11649 | 1 | AE001880 | Deinococc |
| C 43 | 16.4 | 86.3 | 12348 | 1 | AE005856 | Caulobact |
| C 44 | 16.4 | 86.3 | 12891 | 1 | AE005033 | Myobacte |
| C 45 | 16.4 | 86.3 | 15461 | 1 | AE006933 | Myobacte |
| C 46 | 16.4 | 86.3 | 22511 | 1 | AF254925 | Streptomy |
| C 47 | 16.4 | 86.3 | 32394 | 3 | AC022473 | Leishmani |
| C 48 | 16.4 | 86.3 | 36202 | 1 | AB103463 | Streptomy |
| C 49 | 16.4 | 86.3 | 38585 | 3 | AC011913 | Leishmani |
| C 50 | 16.4 | 86.3 | 39428 | 1 | AF322256 | Streptomy |
| C 51 | 16.4 | 86.3 | 39730 | 1 | MTCY8D5 | Myobacteri |
| C 52 | 16.4 | 86.3 | 40570 | 3 | AC073566 | Leishmani |
| C 53 | 16.4 | 86.3 | 40806 | 1 | MSGY140 | Myobacte |
| C 54 | 16.4 | 86.3 | 71623 | 2 | AC087724 | Arabidops |
| C 55 | 16.4 | 86.3 | 85690 | 8 | AB028618 | Arabidops |
| C 56 | 16.4 | 86.3 | 96100 | 8 | AP004670 | Oryza sat |
| C 57 | 16.4 | 86.3 | 99761 | 8 | AC006161 | Arabidops |
| C 58 | 16.4 | 86.3 | 110000 | 2 | LMFLCHR36 | Continuatio |
| C 59 | 16.4 | 86.3 | 110352 | 8 | AC006220 | Arabidops |
| 60 | 16.4 | 86.3 | 113346 | 8 | AC126222 | Oryza sat |
| 61 | 16.4 | 86.3 | 118300 | 8 | AP004165 | Oryza sat |
| C 62 | 16.4 | 86.3 | 119798 | 8 | CNS071P6 | Oryza sat |
| C 63 | 16.4 | 86.3 | 120071 | 8 | CNS08CB3 | Oryza sat |
| 64 | 16.4 | 86.3 | 131278 | 2 | AC108753 | Oryza sat |
| 65 | 16.4 | 86.3 | 141050 | 8 | AC093093 | Oryza sat |


```
C 212 15.8 83.2 1762 3 AY2227001 Anopheles
C 213 15.8 83.2 1762 8 AK067916 Oryza sat
C 214 15.8 83.2 1769 1 STNSIGMAA
C 215 15.8 83.2 1797 6 AK654663 Sequence
C 216 15.8 83.2 1800 1 SCO251005
C 217 15.8 83.2 1845 8 AK108309
C 218 15.8 83.2 1891 8 TAATP2
C 219 15.8 83.2 1909 8 AK070412
C 220 15.8 83.2 1929 6 E05034
C 221 15.8 83.2 1929 8 RICATPB
C 222 15.8 83.2 1964 1 AF237388
C 223 15.8 83.2 1973 8 AK061681
C 224 15.8 83.2 1981 8 AK100872
C 225 15.8 83.2 2000 6 AK655619
C 226 15.8 83.2 2021 8 AK100857
C 227 15.8 83.2 2036 8 D87819
C 228 15.8 83.2 2044 8 HVU272309
C 229 15.8 83.2 2062 8 AK100027
C 230 15.8 83.2 2063 8 AK069323
C 231 15.8 83.2 2118 6 AK653988
C 232 15.8 83.2 2206 8 AK100683
C 233 15.8 83.2 2265 8 AK107091
C 234 15.8 83.2 2271 1 AF217543
C 235 15.8 83.2 2372 8 AK106832
C 236 15.8 83.2 2387 1 RSLGTDGEN
C 237 15.8 83.2 2387 8 AK067392
C 238 15.8 83.2 2471 1 STHTVRO
C 239 15.8 83.2 2471 6 E04368
C 240 15.8 83.2 2480 8 AK119439
C 241 15.8 83.2 2554 1 BBR303064
C 242 15.8 83.2 2554 1 BBR303065
C 243 15.8 83.2 2554 1 BBR303066
C 244 15.8 83.2 2610 1 STVAHPCPA
C 245 15.8 83.2 2633 8 AK072282
C 246 15.8 83.2 2641 8 AK072204
C 247 15.8 83.2 2774 1 AF087458
C 248 15.8 83.2 2866 8 AK072070
C 249 15.8 83.2 2967 8 AK653888
C 250 15.8 83.2 3148 1 SMPFAS2
C 251 15.8 83.2 3248 8 AK071978
C 252 15.8 83.2 3302 8 AF509747
C 253 15.8 83.2 3406 8 AK072337
C 254 15.8 83.2 4003 1 AB010087
C 255 15.8 83.2 4426 8 AF773998
C 256 15.8 83.2 4500 8 AY461703
C 257 15.8 83.2 4585 8 AF509749
C 258 15.8 83.2 4613 8 AF509755
C 259 15.8 83.2 4615 8 AF509756
C 260 15.8 83.2 4615 8 AF509759
C 261 15.8 83.2 4618 8 AF509750
C 262 15.8 83.2 4621 8 AF509757
C 263 15.8 83.2 4622 8 AF509753
C 264 15.8 83.2 4622 8 AF509758
C 265 15.8 83.2 4626 8 AF509751
C 266 15.8 83.2 4627 8 AF509752
C 267 15.8 83.2 4632 8 AF509760
C 268 15.8 83.2 4637 8 AF509754
C 269 15.8 83.2 4649 6 AF362790
C 270 15.8 83.2 4752 1 MTU38939
C 271 15.8 83.2 4857 1 AF143948
C 272 15.8 83.2 5193 1 MSU50335
C 273 15.8 83.2 5373 1 SADNATMS
C 274 15.8 83.2 5721 6 A25910
C 275 15.8 83.2 5811 6 AK616749
C 276 15.8 83.2 6441 6 A14850
C 277 15.8 83.2 6441 6 A19570
C 278 15.8 83.2 6441 6 AK175231
C 279 15.8 83.2 6442 1 BPCYA
C 280 15.8 83.2 6443 6 A07292
C 281 15.8 83.2 6443 6 A32503
C 282 15.8 83.2 6443 6 AR362791
C 283 15.8 83.2 6900 1 PPA250371
C 284 15.8 83.2 7717 8 AF280050
```

```
AY2227001 Anopheles
AK067916 Oryza sat
D17466 Streptomyce
AX654663 Sequence
AJ251005 Streptomy
AK108309 Oryza sat
X74545 T.aestivum
AK070412 Oryza sat
E05034 CNA encodi
D10491 Oryza sativ
AF237388 Azotobact
AK061681 Oryza sat
AK100872 Oryza sat
AK655619 Sequence
AK100857 Oryza sat
D87819 Oryza sativ
AJ272309 Hordeum v
AK100027 Oryza sat
AK069323 Oryza sat
AK653988 Sequence
AK100683 Oryza sat
AK107091 Oryza sat
AF217543 Sinorhizo
AK106832 Oryza sat
Y12481 R.sphaeroid
AK067392 Oryza sat
D13002 S.thermophi
E04368 gDNA encodi
AK119439 Oryza sat
AK072204 Oryza sat
AF087458 Ancylobact
AK072070 Oryza sat
AK653888 Sequence
AF080073 Sinorhizo
AK071978 Oryza sat
AF509747 Hordeum v
AK072337 Oryza sat
AB010087 Pseudomon
AY461703 Chlamydom
AF509749 Hordeum v
AF509755 Hordeum v
AF509756 Hordeum v
AF509759 Hordeum v
AF509750 Hordeum v
AF509757 Hordeum v
AF509753 Hordeum v
AF509758 Hordeum v
AF509751 Hordeum v
AF509752 Hordeum v
AF509760 Hordeum v
AF509754 Hordeum v
AF362790 Sequence
U38939 Mycobacteri
AF143948 Pseudomon
U50335 Mycobacteri
X89899 S.argillace
A25910 B. bronchis
AK616749 Sequence
A14850 B.pertussis
A19570 Sequence 1
AK175231 Sequence
Y00545 Bordetella
A07292 B.pertussis
A32503 B.anthraxis
AR362791 Sequence
AF250371 Pseudomon
AF280050 Oryza sat
```

ALIGNMENTS

RESULT 1

LOCUS AR149059 401 bp DNA linear PAT 08-AUG-2001

DEFINITION Sequence 75 from patent US 6228371.

ACCESSION AR149059

VERSION AR149059.1 GI:15113650

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 401)

AUTHORS Nano.F.E.

TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides

JOURNAL Patent: US 6228371-A 75 08-MAY-2001;

FEATURES

Location/Qualifiers

source

1..401

/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 401;

Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCCGCGATGACC 19

Db 12 GTCGACGCCCGCGATGACC 30

RESULT 2

LOCUS AR340431 401 bp DNA linear PAT 17-AUG-2003

DEFINITION Sequence 75 from patent US 6572865.

ACCESSION AR340431

VERSION AR340431.1 GI:33731979

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 401)

AUTHORS Nano.F.E.

TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides and methods for using same

JOURNAL Patent: US 6572865-A 75 03-JUN-2003;

FEATURES

Location/Qualifiers

source

1..401

/mol_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 19; DB 6; Length 401;

Best Local Similarity 100.0%; Pred. No. 1.2e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0

QY 1 GTCGAGCGCGCGATGACC 19
|||||
Db 12 GTCGAGCGCGCGATGACC 30
|||||

RESULT 3
AE007182/c
LOCUS
DEFINITION
Mycobacterium tuberculosis CDC1551, linear BCT 27-APR-2001
complete genome.
ACCESSION
AE007182 AE000516
VERSION
AE007182.1 GI:13883751
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
1 (bases 1 to 18121)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 18121)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE
Direct Submission
JOURNAL
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..18121
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
207..896
/gene="WT3873"
207..896
/gene="WT3873"
/note="identified by Glimmer2; putative"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK48237.1"
/db_xref="GI:13883752"
/translation="VRSFDSRLTFGIVTYARNWNAANTVRSMDAAGGLHPV
ALMLYEGSPGPGSSWNLVYNLADYAGSPVLLIGYANVDFNWEVRPAGLR
VIGAGYGNPLPGVVAQYTDGSGYSPNLQGPAPFGRCMDNSANGLTPOQFAAAG
VTTCGLPLNLDPEEQTELLTKAREIWDQLRPNAGWPGQUGNEQSDLTTPVDAIAV
IKNDVAALAE"
complement(910..1854)
/gene="WT3874"
complement(910..1854)
/gene="WT3874"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK48238.1"
/db_xref="GI:13883753"
/translation="WPRDNDNSWALTESGATLGVAAAPAAETSDNPLNDPFARI
FVDAAGDSGLWSMYTETLLAGATDLDPLRAPICQMIDFMAARTAFDEYFLATADAG

gene
CDS
1844..2343
/gene="WT3875"
1844..2343
/gene="WT3875"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48239.1"
/db_xref="GI:13883754"
/translation="MGSTPPTPQVFVFAHGOALAGDLDEIVADYADDSFVITPAGI
ARGKEGIOLFVKLLDDIPNALWDLKTIQIFEGDILFLEWTANSVSRVDDGVDTEVFR
DGTVAHTVRYTPHPTK"
complement(2388..2906)
/gene="WT3876"
complement(2388..2906)
/gene="WT3876"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48240.1"
/db_xref="GI:13883755"
/translation="MTPSGTTRNHGYSRIHRPQITVRCRAQKIAQFPFSGESLVEGABQK
VLIDGLSLFORPHEPRIGVGLGRELAFGRHVAQSSGVPFDCRRDPSLVGARVRVDGRE
DGSIVGLVRFQSGDPCSKLLQRHGLILSGRSAGGSACKGCNLTSCAARNAPPGGI
RRQITVPRSDA"
2996..3127
/gene="WT3876.1"
2996..3127
/gene="WT3876.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48241.1"
/db_xref="GI:13883756"
/translation="MNRMTWGCADIGWRYGLEVADPVRRGRHPSDPVEDRALITWR"
complement(3114..3689)
/gene="WT3877"
complement(3114..3689)
/gene="WT3877"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48242.1"
/db_xref="GI:13883757"
/translation="MLSGIQONTLMDNDPLAAGYVADLLVALAVLMLRARRRPE
LARMLLGLTGLVWELPVFGSLAWMTNTEIIEWATPDLPTVTVFLLAHSWDDGLPMT
GMLRALGTGEPAGALGTVLQVWLQTLTALAVELSAIILAGTWSYVDDLLNFWNFMWFR
GHPVTAAOLMTLLAPLCLPFAALVRLRLALTAR"
complement(3795..3977)
/gene="WT3878"
complement(3795..3977)
/gene="WT3878"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48243.1"
/db_xref="GI:13883758"
/translation="MGSTPWCNPNQCITLTPVEVLELAVALRPNPDRTAGAIQRIL
RAQAGDRIALRGSG"
complement(3992..4183)
/gene="WT3879"
complement(3992..4183)
/gene="WT3879"
/note="identified by Glimmer2; putative"

gene
CDS
1844..2343
/gene="WT3875"
1844..2343
/gene="WT3875"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48239.1"
/db_xref="GI:13883754"
/translation="MGSTPPTPQVFVFAHGOALAGDLDEIVADYADDSFVITPAGI
ARGKEGIOLFVKLLDDIPNALWDLKTIQIFEGDILFLEWTANSVSRVDDGVDTEVFR
DGTVAHTVRYTPHPTK"
complement(2388..2906)
/gene="WT3876"
complement(2388..2906)
/gene="WT3876"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48240.1"
/db_xref="GI:13883755"
/translation="MTPSGTTRNHGYSRIHRPQITVRCRAQKIAQFPFSGESLVEGABQK
VLIDGLSLFORPHEPRIGVGLGRELAFGRHVAQSSGVPFDCRRDPSLVGARVRVDGRE
DGSIVGLVRFQSGDPCSKLLQRHGLILSGRSAGGSACKGCNLTSCAARNAPPGGI
RRQITVPRSDA"
2996..3127
/gene="WT3876.1"
2996..3127
/gene="WT3876.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48241.1"
/db_xref="GI:13883756"
/translation="MNRMTWGCADIGWRYGLEVADPVRRGRHPSDPVEDRALITWR"
complement(3114..3689)
/gene="WT3877"
complement(3114..3689)
/gene="WT3877"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48242.1"
/db_xref="GI:13883757"
/translation="MLSGIQONTLMDNDPLAAGYVADLLVALAVLMLRARRRPE
LARMLLGLTGLVWELPVFGSLAWMTNTEIIEWATPDLPTVTVFLLAHSWDDGLPMT
GMLRALGTGEPAGALGTVLQVWLQTLTALAVELSAIILAGTWSYVDDLLNFWNFMWFR
GHPVTAAOLMTLLAPLCLPFAALVRLRLALTAR"
complement(3795..3977)
/gene="WT3878"
complement(3795..3977)
/gene="WT3878"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48243.1"
/db_xref="GI:13883758"
/translation="MGSTPWCNPNQCITLTPVEVLELAVALRPNPDRTAGAIQRIL
RAQAGDRIALRGSG"
complement(3992..4183)
/gene="WT3879"
complement(3992..4183)
/gene="WT3879"
/note="identified by Glimmer2; putative"

/codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48244.1"
 /db_xref="GI:13883759"
 /translation="MRAERARGLFRYQLIREADAHAHKTGKRGKWRRLASREHTD
 PFGKVRISHTIDRWIRN"
 complement (4318. .4644)
 /gene="WT3880"
 complement (4318. .4644)
 /gene="WT3880"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48245.1"
 /db_xref="GI:13883760"
 /translation="MPAPAKALSOVGFRIADLPAPETVGRWLERFAERAEAVRS
 VTVMLRAVDFPVMPPAVGVFAIYAVTVIAVTVIECFALSTVSLATAVAVSGG
 RLVAAPG"
 complement (4779. .4851)
 /gene="tRNA-Arg-4"
 complement (4779. .4851)
 /gene="tRNA-Arg-4"
 /product="tRNA-Arg"
 complement (4882. .4970)
 /gene="tRNA-Ser-3"
 complement (4882. .4970)
 /gene="tRNA-Ser-3"
 /product="tRNA-Ser"
 5048. .6109
 /gene="WT3881"
 5048. .6109
 /gene="WT3881"
 /note="similar to GP:5712092; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="histidinol-phosphate aminotransferase"
 /protein_id="AAK48246.1"
 /db_xref="GI:13883761"
 /translation="MTARLRPELAGLPVYGVKTPVPGAIKLANETVFGPLPSVRAAI
 DRATVNVYDNGCVQLKAALRHLPDPFAPEHVAVGVCSVLCQOLVGVTSVGD
 VVFGRAPELPPQVRVAGAPIQVPLTDHTFDLYAMLATVTRTLIFVCPNPNPTS
 TVFGDALRAVEAPAHILIAIDEAVVEIRDCMRPDSGLVRAHNVVVLSTESKA
 YGLAGLRIGYAGHPDVIATLDRKVIYVFTVSSIGQAAIASLDADELLARTDTVAE
 RARVSAELRAAGFTLPPSQANFVWLPLGSRQDFEQADARIIVRBYTGDRVTVVA
 APEENDAPLRFAPRRWRSDQ"
 complement (6155. .6739)
 /gene="WT3882"
 complement (6155. .6739)
 /gene="WT3882"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK48247.1"
 /db_xref="GI:13883762"
 /translation="MPESRGPDSPTDELACAEALQVLQVLTIGRQDKAKQTP
 CPQDVKKLTSLHNLNVLGGMVGAFLSRADIDSVRLVSGHARSALDAWRHGLE
 GDSVIGPQSMKAVSVFVSFVLVHAWDYAVAVSELKAADSLAEYVLELARKLKP
 EERSVAGNEFVDPEDGGALERLIAFTGRNPAR"
 6763. .7587
 /gene="WT3883"

Query Match 100.0%; Score 19; DB 1; Length 18121;
 Best Local Similarity 100.0%; Pred. No. 9.4e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGACGCCGCGATGACC 19
 Db 14977 GTCGACGCCGCGATGACC 14959

RESULT 4
 MSGV126/c 37164 bp DNA linear BCT 10-DEC-1996
 LOCUS Mycobacterium tuberculosis sequence from clone y126.
 DEFINITION
 ACCESSION AD000012
 VERSION AD000012.1 GI:1717733
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis
 Mycobacterium tuberculosis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.
 1 (bases 1 to 37164)
 Du, L.
 Direct Submission
 TITLE Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
 Reaver Street, Waltham, MA, USA, 02154 du@ctic.com
 COMMENT GSDB:S:1004704
 FEATURES
 Location/Qualifiers
 1..37164
 /organism="Mycobacterium tuberculosis"
 /mol_type="genomic DNA"
 /db_xref="taxon:1773"
 /clone="y126.nt"
 ORIGIN
 Query Match 100.0%; Score 19; DB 1; Length 37164;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGACGCCGCGATGACC 19
 Db 13712 GTCGACGCCGCGATGACC 13694
 RESULT 5
 BX842584/c 244800 bp DNA circular BCT 21-NOV-2003
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 13/13
 DEFINITION BX842584 AL021426 AL022076 AL022120 AL023456 Z80343 Z83864
 ACCESSION Z94121 Z97188
 VERSION BX842584.1 GI:38490370
 KEYWORDS complete genome.
 SOURCE Mycobacterium tuberculosis H37Rv
 ORGANISM Mycobacterium tuberculosis H37Rv
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.
 1
 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
 Tekaita, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
 Barrall, B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 JOURNAL
 MEDLINE 98295987
 PUBMED 9634230
 REFERENCE 2
 Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.
 Re-annotation of the genome sequence of Mycobacterium tuberculosis
 H37Rv
 Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
 JOURNAL
 MEDLINE 22255591
 PUBMED 12368430
 REFERENCE 3
 (bases 1 to 244800)
 Parkhill, J.

| | |
|--------------------|--|
| TITLE JOURNAL | Direct Submission Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On or before Nov 21, 2003 this sequence version replaced gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648, gi:3261687, gi:3261736, gi:3261808. Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/). |
| FEATURES source | 1. .244800 /organism="Mycobacterium tuberculosis H37Rv" /mol_type="Genomic DNA" /strain="H37Rv" /db_xref="taxon:83332" complement(89. .1396) /locus_tag="RV3722c" complement(89. .1396) /locus_tag="RV3722c" /function="UNKNOWN" |
| gene | /note="RV3722c, (MTV025.070c), len: 435 aa. Conserved hypothetical protein, equivalent to O69513 MLCB2407.14 (alias O9CB73 ML2336, 463 aa) HYPOTHETICAL 46.8 KDA PROTEIN from Mycobacterium leprae (426 aa), FASTA scores: opt: 2505, E(): 8.3e-154, (87.25% identity in 424 aa overlap). Also highly similar to O9RUI7 DR1579 CONSERVED HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452 aa), FASTA scores: opt: 1162, E(): 3.1e-67, (44.8% identity in 422 aa overlap); and partially similar to O91371 PA1654 PROBABLE AMINOTRANSFERASE from Pseudomonas aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078, (25.85% identity in 348 aa overlap) and other aminotransferases. Ttparse score is 0.900. N-terminus extended since first submission (previously 408 aa)." /codon_start=1 /evidence=experimental |
| CDS | /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAA18044.2" /db_xref="GI:38490371" |
| gene | /translation="MSFDSLSPQELAAHARHODYAAALQCKMLALDLTRGKPSAEQL LNSQLLSPGDYDPEGTTRNYGGHGLPGLRAIPAELLGIAVFNLIAGNNSLE LMHDIVATSMLYGVDSRPMTQEOGDKFLCPGVDRHFAITETGMIEPIPMIQ DGPVDLIEELVAVDPAIKGMWTVFVCGNPSGVTVSMETVRLVQMRTAAPDRFLPD NAYAVHTLTDPQVDVLGLAAKAGNRPYVFASQSKITFACGGVSPFGSLGNTA WYLQVAGKSGCPDKNQLRHRTFGDAGVRLMLRHQQLAFALVAEVLQRLS ESKIASWTEPKGYFSLDVLPGTARRTVLAKVDGIAVTEAGASFYPRKDPDDKNIR IAPSPSPVDPDLENAVDGLATCALLAATETLLNQGLASSAPNVR" 1613. .1698 /gene="tRNA-Ser (GGA) " 1613. .1698 /gene="tRNA-Ser (GGA) " /product="tRNA-Ser" /note="codon recognized: UCC" /anticodon=(pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" |
| gene | /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11 |
| tRNA | /gene="tRNA-Ser (GGA) " 1613. .1698 /gene="tRNA-Ser (GGA) " /product="tRNA-Ser" /note="codon recognized: UCC" /anticodon=(pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" |
| gene | /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11 |
| CDS | /product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN" /protein_id="CAA18045.1" /db_xref="GI:2960147" /translation="MGRKVAIVLWHSFSGAGLYFYFVLPWPEIMGDTGHSLGTL RIATGALVGLAALPVVFTLLTRKPELGPQALSMRIWSIMAHVLAGALIVGTISGL WMLSDAAGWLFGLYGAAATAVAGFGFVLSFVAELPPPKPKPKPKQRLRR KXTAGDEAPENAEAEATELAAQDEAEVAPESIESPGEGESATREAPAETA TAEPRGGLRNPRTGKTSRRRRIRKSGVQVAKVDE" 2735. .2977 /gene="cut5a" /locus_tag="RV3724A" 2735. .2977 /gene="cut5a" /locus_tag="RV3724A" /EC_number="3.1.1.1." /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)." /note="RV3724A, (MTV025.072), len: 80 aa. Probable cut5a, truncated cutinase precursor (EC 3.1.1.-), similar to N-terminal end of others e.g. O9KX87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 202, E(): 1.5e-06, (56.45% identity in 52 aa overlap); O9XB09 RVD2-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200, E(): 1.5e-06, (61.4% identity in 57 aa overlap); and O00298 CUTI_BOTCI CUTA CUTINASE PRECURSOR from Botrytis cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores: opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap). Also highly similar to others from Mycobacterium tuberculosis e.g. O06318 CUT3 MYCTU RV3451 MT3557 MTCY13E12.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E(): 1.2e-05, (58.0% identity in 50 aa overlap); O50664 CUT2 MYCTU RV2301 MT2358 MTCY339.08c PROBABLE CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E(): 0.0015, (59.2% identity in 49 aa overlap); O06793 RV1758 MTCY28.24 295890 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap); O06319 RV3452 MTY13E12.05, and U00015.11 from Mycobacterium leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase ORF continues as RV3724B CUT5B, frameshifting could occur near position 4169668. Sequence has been checked but no errors found." /codon_start=1 /transl_table=11 /product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A" /protein_id="CAE55632.1" /db_xref="GI:38490372" /translation="MDVIRNARRLAVACTAAVTPGLLSAHVPMSAEPCEPDEVV FARGTEPPGIGSGVGLFDALRFFQWRQVTRGLRR" 2874. .3437 /gene="cut5b" /locus_tag="RV3724B" 2874. .3437 /gene="cut5b" /locus_tag="RV3724B" /EC_number="3.1.1.1." /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)." /note="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b, truncated cutinase (EC 3.1.1.-), similar to C-terminal end of others e.g. O9XB09 RVD2-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 335, E(): 3.4e-12, (53.25% identity in 92 aa overlap); O9KX87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05% identity in 168 aa overlap). Also similar to proteins from Mycobacterium tuberculosis e.g. O06793 RV1758 MTCY28.24 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.5e-29, (57.2% identity in 166 aa overlap). O06319 RV3452 MTCY13E12.05 HYPOTHETICAL 23.1 KDA PROTEIN (226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65% |

identity in 165 aa overlap);
 O06318|CUT3 MYCTULRV3451|MT3557|MTCV13512.04 PROBABLE
 CUTINASE PRECURSOR (247 aa), FASTA scores: Opt: 307, E():
 1.9e-10, (40.7% identity in 167 aa overlap);
 Q10837|CUT1 MYCTULRV1984c|MT2037|MTCV39.35 PROBABLE
 CUTINASE PRECURSOR (217 aa), FASTA scores: Opt: 261, E():
 6.7e-08, (50.9% identity in 169 aa overlap); etc; and
 U00015.11 from Mycobacterium lepra. 5'-end of gene is
 RV3724A|CUT5A; frameshifting may occur near position
 4169668. TBparse score is 0.918."
 /codon_start=1
 /transl_table=11
 /product="PROBABLE CUTINASE [SECOND PART] CUT5B"
 /protein_id="CAE55633.1"
 /db_xref="GI:38490373"
 /translation="WAPGSHLVLAASEDCSSTHCVSQGAKSGLGVAVVNPASNDPAS
 SPPKTVIDIGRIDAGSHISGMASCPQTRVLGVSQGAAGVYTVSAVVPAPVQA
 VPAPMAPEVANHVAATLFGASQAQLGQVGAFFIAIGLYPKTLQLCADSGSDICQD
 GNSPVAGHLVAVNGVMVGQGANFAASRL"

Query Match 100.0%; Score 19; DB 1; Length 244800;
 Best Local Similarity 100.0%; Pred. No. 7.7e-02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
 |||||
 DB 60331 GTCGACGCCGCCGATGACC 60313
 |||||

RESULT 6
 BX248347/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BX248347 278492 bp DNA linear BCT 11-JUN-2003
 Mycobacterium bovis subsp. Bovis AF2122/97 complete genome; segment
 14/14.
 BX248347 BX248333
 BX248347.1 GI:31620456
 complete genome.
 Mycobacterium bovis subsp. Bovis AF2122/97
 Mycobacterium bovis subsp. Bovis AF2122/97
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterinae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.
 1
 Garnier,T., Eighmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
 Pryor,M., Duthcy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
 Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
 Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
 Hewinson,G.
 The complete genome sequence of Mycobacterium bovis
 Online Publication
 PNAS 10.1073/pnas.1130426100 (Microbiology)
 2 (bases 1 to 278492)
 Garnier,T.
 Direct Submission
 Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
 Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
 Mycobacterium bovis sequencing teams, TS Research Group, Veterinary
 Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone,
 Surrey KT15 3NB UK. Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton,Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
 Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
 Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
 rue du Docteur Roux, 75724 Paris Cedex 15, France
 Location/Qualifiers
 1. 278492
 /organism="Mycobacterium bovis subsp. Bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 complement (240..704)
 /locus_tag="MB3713c"
 complement (240..704)

FEATURES
 source

gene

CDS

CDS

2865. .3866
 /locus_tag="Mb3716"
 /note="Mb3716, -, len: 333 aa. Equivalent to Rv3691, len: 333 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 333 aa overlap). Conserved hypothetical protein, similar to O9KYW4|SCE33.19 PUTATIVE SECRETED PROTEIN from Streptomyces coelicolor (387 aa) FASTA scores: opt: 481, E(): 6e-23, (36.6% identity in 358 aa overlap). Equivalent to AAK48160 from Mycobacterium tuberculosis strain CDC1551 (381 aa) but shorter 48 aa." /codon_start=1
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAD955902.1"
 /db_xref="GI:31620460"
 /translation="MAPASTSTGGHALATLGNHGVVVVADSIADVEAARDPSLL LVAQTQYLDNALLAKAPGDLILLVPTSRTRTALTPQLRIAAASFNSQPNCTLR EANRAGSVQMGSDTYQATGDLVLTSCYGGALVFRAGRTITVVVGSNFMFTGGLLP AGNALAMNLAGNRLVWYADHIEGEMSSPSLDLIPENVHMTIWLQVLVLLVA LMKRRIGPLVAELPVTIRASETVEGGRGLYRSRRARDADALRTATLQRLRLG VGAGAPAPAVTTIQRASKADPFVAYHLFGAPATDNDLLQLARALDDIERQVTHS 3863. .4939
 /gene="moxR2"
 /locus_tag="Mb3717"
 /note="Mb3717, moxR2, len: 358 aa. Equivalent to Rv3692, len: 358 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 358 aa overlap). Probable moxR2, methanol dehydrogenase regulatory protein, highly similar (generally longer at N-terminus) to O9KYW3|SCE33.20 PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (329 aa), FASTA scores: opt: 1523, E(): 4.2e-74, (70.9% identity in 330 aa overlap); O9Z538|SC9B2.21c PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (332 aa) FASTA scores: opt: 1008, E(): 1.1e-46, (50.8% identity in 313 aa overlap); O9U267|MOXR-3|PAB0848 METHANOL DEHYDROGENASE REGULATORY PROTEIN from Pyrococcus abyssi (314 aa), FASTA scores: opt: 989, E(): 1.1e-45, (50.65% identity in 302 aa overlap); O9AANI|CC0566 MOXR PROTEIN from Caulobacter crescentus (323 aa), FASTA scores: opt: 988, E(): 1.3e-45, (52.3% identity in 306 aa overlap); ecc. Also similar to O53170|MTV007.26|MOXR|Rv1479 from Mycobacterium tuberculosis (377 aa); and O07392|AF002133.6|MOXR from Mycobacterium avium (309 aa). Also high similarity with several hypothetical bacterial proteins." /codon_start=1
 /transl_table=11
 /product="PUTATIVE METHANOL DEHYDROGENASE TRANSCRIPTIONAL REGULATORY PROTEIN MOXR2"
 /protein_id="CAD95903.1"
 /db_xref="GI:31620461"
 /translation="MTQASNPQAPPTPTGAEPLGYPPQAGGAPTAAPSGPHPRAE AESRDALLAEVAKAVGQDGVISGLVIALLCRGHVLLEGVPGVAKTLIVPMSA ALQLEPKVQVPTDLMGDEVTGSLYDARTAEVFRPGVPTNLLADEINRTPKTC AALLAMBERQVSVGEPEPLNPPIVAATQPIEYGTQYLPPEAQLDFLLKLVTL PARDSEIILDRHAGFDPDLISAINPVAGPAELAAAGAEVHVLANEVLYGIVDIV GATRSFALQGVSPRGATALLGTARSWALSGRDYVTPDDVKAMARPTLRHVMLRP EAELEGATPDGVLGDILASVFPVR" 5073. .6395
 /locus_tag="Mb3718"
 /locus_tag="Mb3718"
 /note="Mb3718, -, len: 440 aa. Equivalent to Rv3693, len: 440 aa (alternative start at 41910), from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 440 aa overlap). Possible conserved membrane protein, similar to O9KYW2|SCE33.21 PUTATIVE LIPOPROTEIN from Streptomyces coelicolor (436 aa), FASTA scores: opt: 875, E(): 3.3e-46, (56.25% identity in 448 aa overlap); O9AAN0|CC0567 HYPOTHETICAL PROTEIN from Caulobacter crescentus (437 aa),

gene

CDS

RESULT 7
 AK072452
 LOCUS
 DEFINITION
 ORYZA SATIVA (JAPONICA CULTIVAR-GROUP) cDNA clone:J023121D13, full insert sequence.
 AK072452 2554 bp mRNA linear PLN 24-JUL-2003
 ORYZA SATIVA (JAPONICA CULTIVAR-GROUP)
 AK072452.1 GI:32982475
 VERSION
 FLI_CDNA; CAP trapper.
 KEYWORDS
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 SOURCE
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.

REFERENCE

AUTHORS

The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otonari, Y., Tsunoda, Y., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodana, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nariakawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kikuchi, S., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

TITLE

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
 Science 301 (5631), 376-379 (2003)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 2554)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiranoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imanura, K., Kagawa, I., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kawai, T., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodana, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Koyada, M., Koya, S., Kuribara, K., Kuroaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariakawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Ooka, H., Ooka, H., Osato, N., Ota, Y., Otonari, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,

gene

CDS

FASTA scores: opt: 355, E(): 2.3e-14, (30.9% identity in 450 aa overlap); P73233|SLR2013 HYPOTHETICAL 48.5 KDA PROTEIN from Synchocystis sp. strain PCC 6803 (435 aa), FASTA scores: opt: 340, E(): 1.9e-13, (29.7% identity in 438 aa overlap); etc. Equivalent to AAK48162 from Mycobacterium tuberculosis strain CDC1551 (475 aa) but shorter 35 aa. Also similar to other hypothetical proteins from Mycobacterium tuberculosis; MTV014_7; MTV007_27; and MTY71_36 M." /codon_start=1
 /transl_table=11
 /product="POSSIBLE CONSERVED MEMBRANE PROTEIN"

Query Match 100.0%; Score 19; DB 1; Length 278492;
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCACCCGCCGATGACC 19

Db 96358 GTCACCCGCCGATGACC 96340

Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shingawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
 Direct Submission
 Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 This clone is one of the 28k full-length cDNA clones from japonica rice.
 URL : http://cdna01.dna.affrc.go.jp/cDNA/

NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
 FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Nariakawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
 Genome Exploration Research Group in Riken Genomic Sciences Center and Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
 Location/Qualifiers
 1..2554
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /culturvar="Nipponbare"
 /db_xref="taxon:3994"
 /clone="J023121D13"

FEATURES

source

Query Match 91.6%; Score 17.4; DB 8; Length 2554;
 Best Local Similarity 94.7%; Pred. No. 4.7e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

1 GTCGACGCGCGGATGACC 19
 |||||
 2256 GTCGACGCGCGGATGACC 2274

RESULT 8

SAMV18862/c

LOCUS SAMV18862 5966 bp DNA linear BCT 03-FEB-2000
 DEFINITION Streptomyces ambofaciens NRRL2240 sar gene and ORF2, ORF3 and ORF4.
 ACCESSION Y18862
 VERSION Y18862.1 GI:6911695

KEYWORDS polyketide synthase; sar gene; Streptomyces ambofaciens regulator; thioesterase

SOURCE

Streptomyces ambofaciens

Streptomyces ambofaciens

Bacteria; Actinobacteridae; Actinobacteriales;

Streptomycetaceae; Streptomyces.

REFERENCE

1

Culebras,E., Martinez,E., Carnero,A. and Malpartida,F.

TITLE Cloning and characterization of a regulatory gene of the SARP family and its flanking region from Streptomyces ambofaciens
 JOURNAL Mol. Gen. Genet. 262 (4-5), 730-737 (1999)
 MEDLINE 20032477
 PUBMED 10628855
 REFERENCE 2 (bases 1 to 5966)
 AUTHORS Malpartida,F.
 DIRECT SUBMISSION Direct Submission
 JOURNAL Submitted (11-MAR-1999) F. Malpartida, Centro Nacional de Biotecnologia, Microbiologia, Lab 218, Campus de la UAM 28049 Canto Blanco, Madrid, 28049, SPAIN
 LOCATION/QUALIFIERS Location/Qualifiers
 1..5966
 /organism="Streptomyces ambofaciens"
 /mol_type="genomic DNA"
 /strain="NRRL2240"
 /db_xref="taxon:1889"
 661..1483
 /gene="sar"
 661..1483
 /gene="sar"
 /note="Protein sequence is in conflict with the conceptual translation"
 /codon_start=2
 /transl_table=11
 /product="Streptomyces ambofaciens regulator (SAR)"
 /protein_id="CAB72140.1"
 /db_xref="GI:6911696"
 /db_xref="GOA:Q9KWX4"
 /db_xref="SPTREMBL:Q9KWX4"
 /translation="MKI OVLGALNAEFNGISVVTAGKPRQILALLALCPGRVVPVT LMEETGTELPQSSMTTQTYTLQRLLGTAMGPDVPSAKDVLATRYGVLQIPAEVADFOYERLVVEGRQAFEDGEDERAALVADLMDPGLVDRVVRVGLLEIEMR MEQSLVTRERRIDAEILGRHVELIAELTDLVARHPQEGHLSQAMVALYSGROAA ALDYVRRLEQRILIELSVPEPQLQELHQLAMLVDPQLDVVAGARRSSTFDLYAA 1842..2822
 /note="ORF2"
 /codon_start=2
 /transl_table=11
 /product="putative thioesterase"
 /protein_id="CAB72141.1"
 /db_xref="GI:6911697"
 /db_xref="GOA:Q9KWX3"
 /db_xref="SPTREMBL:Q9KWX3"
 /translation="VASRTKAPARPARVTLNCLAHAGVAVSYRDNASAVGVDVH ALPLQREARRERPRITTLADLLPGLLTARGPYALYHSGMALVGYTLTRAL ADAGAPPLFLVAGACPPHTTALVDGADLPDDELLPLDITGSLPPGASAPGGLMR RTFLPVLRDDLLAKSLRNAALDPVTGPDVPLVFAAGSDPDLAAPAALRHQQWTT DIELHTVSGGHEFFASSAGLAQHVGACEDHVNAPAGGRR 2619..3890
 /note="ORF3"
 /codon_start=1
 /transl_table=11
 /product="putative polyketide synthase"
 /protein_id="CAB72142.1"
 /db_xref="GI:6911698"
 /db_xref="GOA:Q9KWX2"
 /db_xref="SPTREMBL:Q9KWX2"
 /translation="MKRVVITGIVVAPGVGVTKSLWDLTLTGRATRPTITLFDAGY RSRIAEVDVDFPAAHGFADGTERLDRITQFALVAAREAIADSGTGESVERAPLRCAY SLGSAIGCTTALATVAILSDCGNTWDLDTAADFLYDFVPVPSLAALADCGAQQ PVALVSSGCTSLGDAIGHGADLIREGSDMVAVVGGAEPAPIPTWASFDRLRLTSRN DPATASRPDRTQGFV.GEGAAVLVLEEQSHARPRGAQPYAEILSGLAHSAHMT GLRPGQENAAIRALDQORNLNPADIDIIISAGAGTRNDSHETHAFKAGLGHHAHQ VPVSSIKSMIGHALGAALDITAGALAIRHGIVPTANLHQDPDTCDLDYTFPFARE QRTSTVLVSAGSGFGPHTAAVLTRPQLQAA 3887..5149
 /note="ORF4"
 /codon_start=1
 /transl_table=11
 /product="putative polyketide synthase"
 /protein_id="CAB72143.1"
 /db_xref="GI:6911699"

```
/db_xref="GOA:Q9KWX1"
/db_xref="SPTREMBL:Q9KWX1"
/translation="MSHPAPPAPSPHAPDPGPHPHQALVTGIGITAPNGLGAPAYWD
ALLSGRSLGPTTRDASGYPVRIAGEIITDFATRTNSQOTPAATAAPSPWPRPKES
TTTGAGPQETAGNAGVVTASLAGAEFGQGLAALWSKGATHSVAYOSFASPHAAAP
AOISIRHLRHGHSALITTEAGGIDALARARRINDGAPLWTVGVSTLCPWGAARH
LADGRLSPGEPHHAIRFFAATADGHTVGGEGALLVEDARSARRGATGTVGAGAP
PESTAPRPRURERAEALALAQGLTPAHVVDVADGAGRRAPTAPRAPHCAPCSAPYV
VPTVTKMTGRGLGAGGALDVATLALNAKTVPTTGTTRVAGDCPLDLVTAPRRL
PHLQVALVLAADAEAASTAATVLRAPHG"
ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 5966;
Best Local Similarity 94.7%; Pred. No. 4.4e-03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 4507 GTCGACGCCGCCGATGACC 4489

RESULT 9
BX842638 Neurospora crassa DNA linear PLN 30-NOV-2003
LOCUS Neurospora crassa DNA linkage group I cosmid contig G17B7.
DEFINITION Neurospora crassa DNA linkage group I cosmid contig G17B7.
ACCESSION BX842638
VERSION BX842638.1 GI:38567282
KEYWORDS
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1
AUTHORS Schulte, U., Algn, V., Hoheisel, J., Brandt, P., Fartmann, B.,
Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 60687)
AUTHORS German Neurospora genome project.
JOURNAL Direct Submission
TITLE Submitted (27-NOV-2003) MIPS, Institut fuer Bioinformatik,
GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
G.Mannhaupt@gfz.de Project Coordinator: Ulrich Schulte, Institute
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT this contig is an assembly of cosmid G17B7 from 1 to: 38101, cosmid
library pWocoxX, strain 74-OR-23-1VA, and cosmid 87B11 from 38102
to: 60687, pLORIS6Xh, strain 74-OR-23-1A, cosmid clones are
available at the Fungal Genetic Stock Center, http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
http://www.mwgdna.com
Information on performance of analysis and a more detailed
annotation of this entry and other sequences can be viewed at:
http://mips.gsf.de/proj/neurospora.
FEATURES
source
location/Qualifiers
1..60687
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
/chromosome="I"
1657..3815
/gene="G17B7.010"
complement(join(1657..3038,3143..3539,3810..3815))
/gene="G17B7.010"
complement(join(1657..3038,3143..3539,3810..3815))
/gene="G17B7.010"
/note="match of 17 ESTs
similarity to lysyl-tRNA synthetase (EC 6.1.1.6), Homo
sapiens, SWISSPROT:SYK_HUMAN
contains EST gb:AW710587, AW710588"
/codon_start=1
/product="related to lysyl-tRNA synthetase (lysine-tRNA
ligase)"
```

```
/protein_id="CAE76572.1"
/db_xref="GI:38567283"
/translation="MIFLRPYAMKWPVGARNTAYLRFYHARQHELIOSNALKYPRI
RHEGAPMIPTFREKFKOVEMGKADESVIIGRVQSVRRAGNKLMLDLKGBEYHV
OGLCNFKLOVGTGVELGEFKALAKLLNRGDIVITGRATKPTGELSIEATRLPELLT
PALAPLPKLEBEEVRIQHRHIDMLVNRKTIOTDLRLSHLIXYLRDFFHERGFLBFOQ
PIAENASGATARPFTVTASSTSVNKNLWRIAPELWLRKLVGGVGVKVFELGAFRN
EGDQTHNPEFCMEFYANALDDDLITLIEELICGAASHCRDLIATKLDSLPADMS
QORPQQLFEIPALESALGFKLPDLSASSREDDIALIKHKEIYVGLNKPDITLAK
LIDLNLASELEPYSLTKPLFIHTHPVCMSPLSKSFVCKTGQHVSAETELFNGKELA
NMYEBENDPASOROKFVDOLNAKREASLAPGVYNDEGESAWEVDESYYVALESGLPP
TCGWGAGIERMWMLESGETKEISDCLSFGLNRHVTNVPGARVTGEDEEKARENMDEVEK
KGARERKREKKEKEDMGVSGEA"
complement(1657..3038)
/gene="G17B7.010"
/number=1
complement(3039..3142)
/gene="G17B7.010"
/number=1
complement(3143..3539)
/gene="G17B7.010"
/number=2
complement(3540..3809)
/gene="G17B7.010"
/number=2
complement(3810..3815)
/gene="G17B7.010"
/number=3
4460..4985
/gene="G17B7.020"
join(4460..4551,4805..4985)
/gene="G17B7.020"
/note="identical to TREMBL:AF343076.1
contains EST gb:AW714171, AW709501, AW712102, AW708270,
AW711482, AI329860"
/codon_start=1
/product="mitochondrial import inner membrane translocase
subunit TIM10"
/protein_id="CAE76573.1"
/db_xref="GI:38567284"
/translation="WFGIGRPOPTSAEKIAVAENELKVAEMHSMVVICILKCIDKS
YREGDUSKGSVCLDRCAAKFFETHQKISDQLQKETAQGGGGFGM"
4460..4551
/gene="G17B7.020"
/number=1
4552..4804
/gene="G17B7.020"
/number=1
4805..4985
/gene="G17B7.020"
/number=2
5985..9947
/gene="G17B7.030"
complement(5985..9947)
/gene="G17B7.030"
complement(5985..9947)
/notes="contains EST gb:BG280456, BG280455"
/codon_start=1
/product="putative protein"
/protein_id="CAE76574.1"
/db_xref="GI:38567285"
/translation="MTDPSNLSNLRPRRTATKQPDLDSDRHDYASTAISRQDHQVTT
NLPDVSVTSDNDLVKDYDDASIPDQRLPSRHPDRADNRFSGALRAHKNRNGA
FLPSALGATQHTQLOLPKRQSLLEHGRNRITKNGMDRSPGRNKLTPGNEVLGL
VORSGTTPSVRQTSINDSAVTNRRSGSAGETIIGTNPSPRTSMALLMESAIHV
MALNLSERGRASIRRVISQVPPRGPENSTGSLRQHLQOQSKVRNMSRPDRHS
PRIASABSLQASAGLEAYQVHESQSTLARAQREYGLMACYRELLLELLPLKAPPS
ATYTSVSPLETFNGFNFTPTQDAETIGRPYNPLOYIENRKIRABERTKLDGEGQ
GNDVBERSEWDEVAKMIAGKQDRIFGNPSIPFPSPTRILUSCEIPFQSLGSKPTAI
PKKSGTIFNSNITADLADIYWLQDDNKLEVEDNRNWRKLPFQPGAGIYSLSRDIT
POSSKIFNSNITADLADIYWLQDDNKLEVEDNRNWRKLPFQPGAGIYSLSRDIT
ATRESHRANSIYNRDLIRHRRSFSSSDTSDSRKGRGRHGHTISSSGKDIARQME
```

DMIAEQORD:EAHPLVDQDQOQLPDAGLAMSEKQGEWPRITSTSDGADNVEGRPPA
SRPAPADEGNTKYQAPAPAGVLAGRAGVSGDGRISDYDTSRSPYSPDLRPVH
DALIPRESTNSKSRSPPIARSKNPLTKVRSIFRDSNEHLAELPSHIGRDN
PESTQFUDQRPVPTLLGFGTGMSPPKRKTSPRLKIVTKGTDSSQKNGKSTGI
YKJRDGNSLKGFLFGORQDVLRSQSVKVSDFWKEAPGSPDXVSSDESDAES
RGRSLRPPSPKVSRSNHEKTKNGKYNLDVMPQFVPTAEKHHSKSKSVGSDIP
NGLSASPPQSPRSRSPFELLKPKIDIONASPTSSPGFLEITRALRFGWSDVDPV
EGKRSKTDVGVGDADARLNLMSKMGQRLVNSASAGWSIAADRENNATVPVP
QTVISKEIARLALLSSGIIHAMEIDRAKADKLLNANSTDDENANTTTASSHINN
NSRLSPKDIIVALLSPDPSRTOQLLVPIAQDILFPAANILSTIQNSATAFQSSVAF
QTOTAPMTNVRVREALRSISSELTETRRAAEADAGGLMTHORLKVVRVYDMIDR
MVRRRRRFRFVRAGWLAWEVLVGFVWFVFWMIARVLLGVGRVGVGWEWLLWL

complement(5985..9947)
/gene="G17B7.030"
/number=1
11752..13517
/gene="G17B7.040"
complement(join(11752..11811,11907..13517))
/gene="G17B7.040"
complement(join(11752..11811,11907..13517))
/gene="G17B7.040"
/note="similarity to unknown protein AT3g51130,
Arabidopsis thaliana, TREMBL:AY037194.1"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAE76575.1"
/db_xref="GI:38567286"
/translation="MSTPTSPSPFLSQAQYPRGALGVLAGSLHDLTLRLKTEP
QKRDNLVKTSPTEPTKSIIVIDLIPANGIRLOFQDAPEQRLRIEIVDFTKNHFIRKAA
DNFEDLVRSASASTPSTNDNPAFTFRIHQRLGTYDGEFIDGPKTPDQAVDQGL
VYLSWPGVAFSGKMSYSAKKNVDLPSFSTVETMIFSGDSMTQAROTLWQVL
PSLTKFTPLKGVFDEYSLIKIYAGKLOVFRKTDQHFWRIGQTTQPDVLVSHL
GPPDAIRKNDHNRFAKVRGSDALRPSADLKHQDDITDIDQSSANAGSGYRSS
DDETSVDSVGHVSGECFYFNGFDVLVSKTAPSPVPPQEQKNVNPIDRWLL
TANASIVATKIVLHNGVGSYFPHRRRCWEIPYLSASDVINSEDFPQIBEL
RKQWQSTYDGLKQAGVNLNKGWSDSPGSSIEFLGWEGDGAAPADAGTAKK
FDGPADSTTLYGPGHVFVEVLKNGAVTALVLRDTLLIGGAGYMWVNG"
complement(11752..11811)
/gene="G17B7.040"
/number=1
complement(11812..11906)
/gene="G17B7.040"
/number=1
complement(11907..13517)
/gene="G17B7.040"
/number=2
14789..18540
/gene="G17B7.050"
join(14789..15271,15322..17478,17536..18540)
/gene="G17B7.050"
/note="strong similarity to kinesin-related protein bimC,
Emicella nidulans, PIR:A34795"
/codon_start=1
/product="probable kinesin-related protein bimC"
/protein_id="CAE76576.1"
/db_xref="GI:38567287"
/translation="MERKSPMSRPPSVAGTRSNSSMRPPTRTNTASSTGROSALRN
ASAGSGRAESGAASPADQSVATAGAKERDFDPEDRGTTNNVVRCGRNDRE
VKNSAVYVTEGKGLVLSMGNVSNKTYTFDFESQADQVADVNVKPKILE
EMLAGYNTFYAGTGTGTYTMSGDMQTFPMPLTAGIIPRLVHALNKLEVDK
EHCVRGFIEYNEELDLAEDNTKLIPDNKKGHATTIVQGEERHLSAADG
LWQEGSVRRQVATKNDLSSRSTVFTIVHKOQSTENGEDLMGKXNLVLDLAG
SENIRGAENKRAAEAGLNLKSLITGRVINADVSPHPIYRESKTLRLQDSLGG
RTKCIATISPASNNVEETISLDYAFRAKRNKQINLKNTLLREFATEIEK
LKSLLIATGRQVYLNDAVEILTQVNSRRILTEQAAARIETLENLNKQELYT
MSTFMGLKREHATKQDQDKVLPOTELVLAATKSLAEAHILRAHQTTEOKLA
KGBELISTVKTQVDEGLHAKNRSDHLNRSWMSQSVQGITELVESRTAE
FKQGEHSISVSERMOSFVQEDLKLSTQAFILQDNLHFAERTQLLEKQKSKEE
MDNVEIKVVRDITVKQGVESQAIASAEIRAGDVLGELTTFHSQHLASYSLGKE
CKSIFGDLLOTSIQAKAEARLQELCASQIVESNATVSRITQVLEERQAAEE
RONLLQIGSLINSQAEQMSRLAKGTALIROAVMDSNATFEGSMSYMDGMWAMNEK
DSRLMEDVVKSRDVLNKKMKDDWATADEHSNSIQNTKSVHAETVRVVDQAQLDLDLQ

MODLDQFVSRRARSNSVHHEQHVSVSMENLSTVKSFINIADHGERDRDRDLSGEM
DNDAKILRESLEPDESRLQALAREDIQSTKREYPTGDTPEKKRYEYPTDLPRIT
AGHETLIAGMDQKPSITTTTSPAKRATMTVIVSVSPSRSPKSVLPVSPVSPRNP
FCTPGPATSTRNKSHMPDPSLRVNPVNLGGGTSVFGELSSSTMSMAPSTSVIGFHAN

Query Match 91.6%; Score 17.4; DB 8; Length 60687;
Best Local Similarity 94.7%; Pred. No. 3.7e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 2647 GTCGACGCCGCCGATGACC 2665

RESULT 10
AC068950/c
LOCUS
DEFINITION
Oryza sativa chromosome 10 BAC OSJNBa0041P03 genomic sequence,
complete sequence.
AC068950
AC068950.11 GI:22128701
HTG.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzeae; Oryza.
1 (bases 1 to 139614)
Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Gansberger, K., Kim, M.M.,
Overton II, L.L., Bera, J.J., Tsitrin, T., Krol, M.I., Jarrahi, B.,
Jin, S.S., Koo, H., Zismann, V., Hsiao, J., Blunt, S., Vanaken, S.S.,
Uterback, T.T., Feldblum, T.V., Yang, Q.Q., Haas, B.J., Suh, B.B.,
Peterson, J.J., Quackenbush, J., White, O., Salzberg, S.L. and
Fraser, C.M.
Oryza sativa chromosome 10 BAC OSJNBa0041P03 genomic sequence
Unpublished
2 (bases 1 to 139614)
Buell, R.
Direct Submission
Submitted (13-MAY-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 139614)
Buell, R.
Direct Submission
Submitted (31-JUL-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
4 (bases 1 to 139614)
Buell, R.
Direct Submission
Submitted (07-AUG-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
On Aug 7, 2002 this sequence version replaced gi:22024565.
Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0041P03 is from Oryza sativa chromosome 10
The orientation of the sequence is from Sp6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Fgenesh (<http://www.softberry.com/>),
GENSCAN+ (Chris Burge,
<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkMM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer
(Mihaila Partea and Steven Salzberg, contact.mpartea@tigr.org),
searches of the complete sequence against a peptide database and
the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as unknown proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as hypothetical proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,

<http://genome.wustl.edu/eddy/ERNAscan-SE/>. Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

This BAC overlaps with rice BAC OSJNBa0051D19 (AC023240) and OSJNBa0066108 (AC092489).

There is a small sequencing gap (<500 bp) in this BAC. Closure effort continues.

FEATURES

Source

1. 139614

Location/Qualifiers

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="10"

/map="R2303"

/clone="OSJNBa0041P03"

/note="japonica cultivar-group"

371..8253

/gene="OSJNBa0041P03.1"

/note="similar to second largest subunit of RNA polymerase I GB:AC52850 GI:1621594 (Mus musculus), EST AU184056 from this gene"

Join(<371..444,537..654,1422..1506,2098..2190,2722..2834,

3721..3848,3924..4003,4093..4282,4424..4522,4664..4828,

4951..5025,5415..5518,5601..5676,6300..6422,6532..6675,

6989..7201,7373..7606,7810..8253)

/gene="OSJNBa0041P03.1"

Join(<371..444,537..654,1422..1506,2098..2190,2722..2834,

3721..3848,3924..4003,4093..4282,4424..4522,4664..4828,

4951..5025,5415..5518,5601..5676,6300..6422,6532..6675,

6989..7201,7373..7606,7810..8253)

/gene="OSJNBa0041P03.1"

/codon_start=1

/product="putative DNA-directed RNA polymerase I subunit,

5'-partial"

/protein_id="AA092814.1"

/db_xref="GI:22128702"

/translation="VAEAVLKDYIFVHLQNHDKFNLLIFMLQKLYAIVDQTSAPDKA

DALQYVELLPGLHITVFLKDLQWLRKLLVEEATKNSFDLDFDFLEGGV

RGTSVETSSGKPCMLGAEQVRKFLSXTAYGKAIQSMIKVGKVASQGLDLP

QRDQVTHAERLNFHYISHSFVSRRGSSFAKMTTSVRKLLPESWGLPCVHTPDE

CLLHNTSTCRISSTFNSGATKDFOKIKWSLIARLVGMAQLPRTERTGPPEV

LGVHDGTCIVGSIASAKIEEVIPEDLGVYVPLSHGGAYPGLYLFTWPARLPARPE

TRCPDGGGRNKLFPAPHEIHTAILSVVANTPWSHDNOSPRNNVQCMAKQTMG

FCGQALKRVDKFLHQTPTPIVATYISKYCMDFPSGTNAIVAVLSTGYDMED

AMLNKASRDVMFRGHIFQTEIDLKSAKSDNVTFFCKNSRDTTAAISDGLPR

IGENITFPEQYYSVNNLTGTVPRIKLKSGSPAIDYVAVNGTNRFDRLQKANTLR

VNRTPIGKFSRSGHQGSQLWPDIDMFPSTANTGRPDLIINPHAFPSMTIAMLL

ESIAAGSLGKKTIDPFPASVYKERSNSIVDELGMLASGFNYHGTILYSGVFG

TEMKCEIFLGPVYTRLHMYSDKQVTRTTCRIQITROPIGRGRGGIRFGEMED

ALLAHGASVLLDLHSCSDYHADVCSIGSLTATVIKSESOKKAKRMGLIPTVK

FPKPACQACKTSKGMETVAMPYVFRYLASELANMLKLELRLSNRTEHTTTTSES"

4861..4900

/rpt_family="tandem"

6064..6065

/gene="OSJNBa0041P03.1"

/note="about 779 bp of E. coli insertion sequence is1E

sequences deleted"

complement(9020..9102)

/rpt_family="(CGG)n"

9020..9095

/rpt_family="tandem"

9046..9102

/rpt_family="tandem"

9109..13951

/gene="OSJNBa0041P03.9"

/note="predicted by genemark-ES"

Join(<9109..9130,9219..9343,10810..10881,10978..11487,

11595..11909,12003..12059,12235..12561,12641..12847,

12925..13011,13097..13191,13410..13544,13630..13951)

/gene="OSJNBa0041P03.9"

CDS

Join(9109..9130,9219..9343,10810..10881,10978..11487,

11595..11909,12003..12059,12235..12561,12641..12847,

12925..13011,13097..13191,13410..13544,13630..13951)

/gene="OSJNBa0041P03.9"

/codon_start=1

/product="hypothetical protein"

/protein_id="AA092816.1"

/db_xref="GI:22128704"

/translation="MSDSDEYVDLPVSDDEEEDSEDEKVGSRKKAKVHAKQLK

RLQEKDPFYKYLECDKELFFDDDDFDNNGSAEKHSVPKPKVPIKPIWQMV

DMWCOGADKIGSIRSILEARAKACHYGESENNAPKFSVWSSGLDKVHFLVKN

MRIARLLELDAPSGFGKKTYSSELMITQWKRHGRMLRLYLVAHMLTILDEQMA

FIVHRVASAVFLAFAFALLRIYVALLHWSRGRLAMPVSLFLRLDICIQLGSECL

DTSLGIYKAYLVNCKLSKISQKLIQIFLGNVCVRLYNNVDPOYAHAFVFIQL

AVILRGALTEGPKTSKDKKQKESIKPTKKRMEKSYQKYDQWYIFCLELMTSVVCGC

SSEEDRLPLAYPLTIQIHGVACLVPSARYFPVLRVCMRLNRIEATGTPIPVSSLLI

DMLEKELCGKPDVAKVNLFSVKQDKTKVTRAFQACIFSAVDFSLAKHLAOWSY

SIATFEMSPLTLVRLQNFCKTVKADREFRREIKDLIHQIKASAEFVSSKRAIGFSPND

PAVDSFLQVEKAKSPSLSKYVATILQRSQDMSLDLDTSYVGAESSTFSERLSEAD

KKQDFQDDGEDIATFSKNLLTEKKTKTPKSKRAHNDHDAVEEDIVEDLLISSD

EDEDDEKPMSEDDGSGNVEDESDDDFVDPDSQWKKQKKEKSKRNRPSPSKGSS

TTKRKEKIPHPKKAKH"

9235..9284

/rpt_family="(GGA)n"

9522..9579

/rpt_family="tandem"

9539..9570

/rpt_family="tandem"

14553..14605

/rpt_family="tandem"

14628..14666

/rpt_family="tandem"

complement(14849..14883)

/rpt_family="AT-rich"

15203..16355

/gene="OSJNBa0041P03.11"

/note="predicted by genscan"

Join(<15203..15446,15511..>16355)

/gene="OSJNBa0041P03.11"

Join(15203..15446,15511..16355)

/gene="OSJNBa0041P03.11"

/codon_start=1

/product="hypothetical protein"

/protein_id="AA092818.1"

/db_xref="GI:22128706"

/translation="MKSPPRAASAPAPAPMPMLSPAGKPKHFTLQDSSAERDA

DLEAFWRARLVVSSSRVAGDLAYLRRRSRSPARGRGGGGGGGAQVYCGRGG

VVAHVAGRVAGGVDVRRGVAGGGGQACGRDPAGVRRALRCRRRAEGGGGRS

RWRRDGVHVARRRGAGGPDAGVGAGVLLRRRRRREAGAVLDGGEAGEPGLRRAR

RAGGPRAAARPPFVAGRGAGEGLLRKRGVPEFGAARACRRRRRGDAAAEAGG

DLQGTAEASGCGAPVAAAKLREVEALAGAEHDNGMYVMAVPANGELAEPLGLQ

ORLVQTPLENILOQGTGDDGAPAGG"

complement(15221..15256)

/rpt_family="GC-rich"

15517..15661

/rpt_family="(CGG)n"

complement(15627..15793)

/rpt_family="(CCGCG)n"

16346..16374

/note="Region with a single high quality sequence

confirmed with a second sequence."

17268..19522

/gene="OSJNBa0041P03.10"

/note="predicted by Igenesh"

Join(<17268..17292,17417..17767,17825..18969,

19490..>19522)

/gene="OSJNBa0041P03.10"

Join(17268..17292,17417..17767,17825..18969,19490..19522)

/gene="OSJNBa0041P03.10"

/codon_start=1

Query Match 91.6%; Score 17.4; DB 8; Length 139614;
Best Local Similarity 94.7%; Pred. No. 3.5e+03;

| | |
|--|---|
| Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
| Qy | 1 GTGACCGCCCGCATGACC 19 |
| Db | 76086 GTGACCGCCCGCATGACC 76068 |
| RESULT 11 | |
| LOCUS | AP005494 |
| DEFINITION | Oryza sativa (japonica cultivar-group) chromosome 8 clone OSUNBa0086M15, *** SEQUENCING IN PROGRESS ***. |
| ACCESSION | AP005494 |
| VERSION | AP005494.1 GI:21699014 |
| KEYWORDS | HTG; HTGS PHASE2. |
| SOURCE | Oryza sativa (japonica cultivar-group) |
| ORGANISM | Oryza sativa (japonica cultivar-group) |
| REFERENCE | |
| AUTHORS | Sasaki, T., Matsumoto, T. and Katayose, Y. |
| TITLE | Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC clone-OSUNBa0086M15 |
| JOURNAL | Published Only in Database (2002) |
| REFERENCE | 2 (bases 1 to 154288) |
| AUTHORS | Sasaki, T., Matsumoto, T. and Katayose, Y. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (04-JUL-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468) |
| COMMENT | |
| NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved. | |
| * NOTE: This is a 'working draft' sequence. | |
| * by the finished sequence as soon as it is available and the accession number will be preserved. | |
| FEATURES | |
| source | location/Qualifiers 1..154288 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="8" /clone="OSUNBa0086M15" |
| ORIGIN | |
| Query Match 91.6%; Score 17.4; DB 2; Length 154288; Best Local Similarity 94.7%; Pred. No. 3.4e+03; | |
| Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
| Qy | 1 GTGACCGCCCGCATGACC 19 |
| Db | 52622 GTGACCGCCCGCATGACC 52640 |
| RESULT 12 | |
| LOCUS | AF232689 |
| DEFINITION | Rat cytomegalovirus Maastricht, complete genome. |
| ACCESSION | AF232689 AF046125 U50550 AF077758 U91788 AF133339 U57441 U57442 U87867 AF136519 |
| VERSION | AF232689.2 GI:20178619 |
| KEYWORDS | |
| SOURCE | Rat cytomegalovirus Maastricht |
| ORGANISM | Rat cytomegalovirus Maastricht Viruses; dsDNA viruses, no RNA stage; Herpesviridae; |
| REFERENCE | |
| AUTHORS | Vink, C., Beuken, E. and Bruggeman, C.A. |
| TITLE | Complete DNA sequence of the rat cytomegalovirus genome |
| JOURNAL | J. Virol. 74 (16), 7656-7665 (2000) |
| REFERENCE | 20366325 |
| AUTHORS | 10906222 |
| TITLE | 10 (bases 1 to 230138) |
| JOURNAL | Grujthuisen, Y.K., Beuken, E., Bruggeman, C.A. and Vink, C. |
| REFERENCE | Rat cytomegalovirus R89 is a highly conserved gene which expresses a spliced transcript |
| TITLE | Virus Res. 69 (2), 119-130 (2000) |
| REFERENCE | |
| AUTHORS | Beuher, P.S., Vink, C., Van Dam, J.G., Grauls, G., Vanherle, S.J. and Bruggeman, C.A. |
| TITLE | The R33 G protein-coupled receptor gene of rat cytomegalovirus plays an essential role in the pathogenesis of viral infection |
| JOURNAL | J. Virol. 72 (3), 2352-2363 (1998) |
| REFERENCE | 98139136 |
| AUTHORS | 9499096 |
| TITLE | 5 (bases 168213 to 176656) |
| JOURNAL | Beisser, P.S., Kaptein, S.J., Beuken, E., Bruggeman, C.A. and Vink, C. |
| REFERENCE | The Maastricht strain and England strain of rat cytomegalovirus represent different betaherpesvirus species rather than strains |
| TITLE | Virology 246 (2), 341-351 (1998) |
| JOURNAL | 98327782 |
| REFERENCE | 9657952 |
| AUTHORS | 6 (bases 98664 to 102381) |
| TITLE | Beisser, P.S., Grauls, G., Bruggeman, C.A. and Vink, C. |
| JOURNAL | Deletion of the R78 G protein-coupled receptor gene from rat cytomegalovirus results in an attenuated, syncytium-inducing mutant strain |
| REFERENCE | J. Virol. 73 (9), 7218-7230 (1999) |
| JOURNAL | 99370163 |
| REFERENCE | 10438809 |
| AUTHORS | 7 (bases 23514 to 27145) |
| TITLE | Beuken, E., Grauls, G., Bruggeman, C.A. and Vink, C. |
| JOURNAL | The rat cytomegalovirus R32 gene encodes a virion-associated protein that elicits a strong humoral immune response in infected rats |
| REFERENCE | J. Gen. Virol. 80 (Pt 10), 2719-2728 (1999) |
| JOURNAL | 20037837 |
| REFERENCE | 10573166 |
| AUTHORS | 8 (bases 194897 to 204137) |
| TITLE | Beisser, P.S., Kloover, J.S., Grauls, G.E., Blok, M.J., Bruggeman, C.A. and Vink, C. |
| JOURNAL | The R144 major histocompatibility complex class I-like gene of rat cytomegalovirus is dispensable for both acute and long-term infection in the immunocompromised host |
| REFERENCE | J. Virol. 74 (2), 1045-1050 (2000) |
| JOURNAL | 20091365 |
| REFERENCE | 10623772 |
| AUTHORS | 9 (bases 1 to 230138) |
| TITLE | Vink, C., Beuken, E. and Bruggeman, C.A. |
| JOURNAL | Complete DNA sequence of the rat cytomegalovirus genome |
| REFERENCE | J. Virol. 74 (16), 7656-7665 (2000) |
| JOURNAL | 20366325 |
| REFERENCE | 10906222 |
| AUTHORS | 10 (bases 1 to 230138) |
| TITLE | Grujthuisen, Y.K., Beuken, E., Bruggeman, C.A. and Vink, C. |
| JOURNAL | Rat cytomegalovirus R89 is a highly conserved gene which expresses a spliced transcript |
| REFERENCE | Virus Res. 69 (2), 119-130 (2000) |
| JOURNAL | |

```

MEDLINE
PUBMED
REFERENCE
11 (bases 61631 to 74978)
AUTHORS
Beuken,E., Slobbe,R., Bruggeman,C.A. and Vink,C.
TITLE
Direct Submission
JOURNAL
Submitted (05-MAR-1996) University of Limburg, Medical
Microbiology, P.O. Box 5800, Maastricht 6222 WH, Netherlands
12 (bases 1 to 1511)
AUTHORS
Vink,C.
TITLE
Direct Submission
JOURNAL
Submitted (03-MAY-1996) Medical Microbiology, University of
Limburg, P.O. Box 5800, Maastricht, AZ 6202, Netherlands
13 (bases 229050 to 230137)
AUTHORS
Vink,C.
TITLE
Direct Submission
JOURNAL
Submitted (03-MAY-1996) Medical Microbiology, University of
Limburg, P.O. Box 5800, Maastricht, AZ 6202, Netherlands
14 (bases 74973 to 78970)
AUTHORS
Vink,C., Beuken,E. and Bruggeman,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (30-JAN-1997) Medical Microbiology, Maastricht
University, P.O. Box 5800, Maastricht, AZ 6202, Netherlands
15 (bases 25540 to 27709)
AUTHORS
Beisser,P.B., Vink,C. and Bruggeman,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (03-MAR-1997) Medical Microbiology, University of
Maastricht, P.O. Box 5800, Maastricht, AZ 6202, Netherlands
16 (bases 168213 to 176556)
AUTHORS
Beisser,P.S., Kaptien,S.J.F., Beuken,E., Bruggeman,C.A. and Vink,C.
TITLE
Direct Submission
JOURNAL
Submitted (03-FEB-1998) Medical Microbiology, Maastricht
University, P. Debyeilaan 25, Maastricht, AZ 6202, Netherlands
17 (bases 98664 to 102381)
AUTHORS
Beisser,P.S., Vink,C., Beuken,E. and Bruggeman,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (13-JUL-1998) Medical Microbiology, Maastricht
University, P. Debyeilaan 25, Maastricht, AZ 6202, Netherlands
18 (bases 194897 to 204137)
AUTHORS
Beisser,P.S., Grauls,G., Bruggeman,C.A. and Vink,C.
TITLE
Direct Submission
JOURNAL
Submitted (07-MAR-1999) Medical Microbiology, Maastricht
University, P. Debyeilaan 25, Maastricht, AZ 6202, Netherlands
19 (bases 23514 to 27709)
AUTHORS
Vink,C., Beuken,E. and Bruggeman,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (18-AUG-1999) Medical Microbiology, University of
Maastricht, P.O. Box 5800, Maastricht, AZ 6202, Netherlands
20 (bases 1 to 230138)
AUTHORS
Vink,C., Beuken,E. and Bruggeman,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (04-FEB-2000) Medical Microbiology, University of
Maastricht, P.O. Box 5800, Maastricht 6202 AZ, Netherlands
21 (bases 1 to 230138)
AUTHORS
Vink,C., Beuken,E. and Bruggeman,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (18-APR-2002) Medical Microbiology, University of
Maastricht, P.O. Box 5800, Maastricht 6202 AZ, Netherlands
REMARK
Sequence update by submitter
COMMENT
On Apr 18, 2002 this sequence version replaced gi:9800235.
FEATURES
source
1..230138
/organism="Rat cytomegalovirus Maastricht"
/mol_type="genomic DNA"
/strain="Maastricht"
/db_xref="taxon:97900"
1..505
/note="terminal left"
/evidence="experimental"
complement(517..1173)
/gene="r1"
complement(517..1173)
/gene="r1"
/codon_start=1
repeat_region
gene
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
CDS
Genome sequence of an industrial microorganism Streptomyces
avermittilis: deducing the ability of producing secondary

```


metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
11572948
2
Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,
Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
Complete genome sequence and comparative analysis of the industrial
microorganism *Streptomyces avermitilis*
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
12692562
3 (bases 1 to 299175)
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,
Sakaki, Y. and Hattori, M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0065, Japan
(E-mail: biocentre.go.jp, URL: <http://www.bio.nite.go.jp/>,
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda (*1), Jun
Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Mayumi
Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi
Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi
Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7)
and Satoshi Omura (*1, *3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
<http://avermitilis.ls.kitasato-u.ac.jp>.
Location/Qualifiers
1..299175
/organism="Streptomyces avermitilis MA-4680"
/mol_type="Genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/note="This strain is also named as strain: ATCC 31267,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermitinus"
complement(155..550)
/note="SAV445"
Tn4561, truncated
/codon_start=1
/transl_table=1
/product="putative transposase"
/protein_id="BAC68155.1"
/db_xref="GI:29604084"
/translation="MAGPPRSATRSRTADREDSAPADPADPFGYRRQIKVQANLQ
EGHALLARRAGQLYQYQDMQIGALGLVNLALFNTRYMDAALTQLRADGFEVR
DEDVARLSPFWRHHINVLGRYSFQLPDL"
complement(474..740)
/note="IS630 family, truncated
SAV446"
/codon_start=1
/transl_table=1
/product="putative transposase"
/protein_id="BAC68156.1"
/db_xref="GI:29604085"
/translation="MPTQNAQPSSTVTAARVSTAGSVYSNAPTSFVMPAPSPFGRRR
EPGSLGAVRAYDIVRMILSRDGRPTPLGDAITYGSRRLCTSCA"
766..987
/note="SAV447"
/codon_start=1

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
CDS
CDS
CDS

/transl_table=11
/product="putative transposase"
/protein_id="BAC68157.1"
/db_xref="GI:29604086"
/translation="MUVVAVGCAITSVPTAGGVVAEPVCVRRITDQGGKQLQOIVRRG
SASSVPRPRAMLLASAGNRPVIAQLV"
complement(1954..2571)
/note="SAV448"
/codon_start=1
/transl_table=11
/product="putative Tetr-family transcriptional regulator"
/protein_id="BAC68158.1"
/db_xref="GI:29604087"
/translation="MFTSRRARERANTRERIEAALHVLTEGIAALIRRIATQVE
YSAPVYQHFANKDALVLVBHAGHRLMISEFSQAAQEPDTRRMTRIASQVYRFAGE
HPLFQVMDNDPVDADERVAEPVIDVLKELITSWAADHVVLADFDQACEILWGTL
YGIASLYLGSVGNDRARYLAEQALHTILLGWRTETPANGRPASS"
2694..3098
/note="SAV449"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68159.1"
/db_xref="GI:29604088"
/translation="MIYHNRVTMKATATPEQIEGVLESWRNQRNPAIKSFVQGRN
HGGDYTYGAVFVVEHLDGLFAYLTHPTTYQTDHGLHLVERLEIFDVSDDDPDLNAK
IQELHRRNELNPQIAGMLADVPTTYSGVDD"
3127..4029
/note="SAV450"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68160.1"
/db_xref="GI:29604089"
/translation="MAKHVIGAGSIGTNNVARLLAEGGESVRIIVTSGSGPHEPLVDR
VADADSPRLTSLSRGAGVIYHCANPPSYWRELLPLOTAAITAAKANDAVLALT
GSLYAYGPQGGRMNEHTPMATGHKRLRRMWEQALAGGIRTVRSGSDYIKGDAN
GIYLLIKSALQKGAAWITGHLDMPHTTFNGDQAALITLASEERAGRPWHVPSA
PAVTIRELARRYAVAAGRPPIKLIQIPRSVTRTAGLIVFIAREMAEMDYQWYAPFMD
AHTETTFGLTDLTDVARNVG"
complement(4130..4498)
/gene="bidg"
complement(4130..4498)
/gene="bidg"
/note="SAV451"
/codon_start=1
/transl_table=11
/product="putative anti-anti-sigma factor"
/protein_id="BAC68161.1"
/db_xref="GI:29604090"
/translation="MTDAFHIDVPHPTDAALAVIALSGEPDIAAPAVRARALELIANG
HPDLVADLSGVTFCSSGLGALVGIWRCAKDADGSLTLAAIPDLRLSRLLSVTGMDAFL
PAYSSADAALAARQGNRTTA"
4635..5156
/note="SAV452"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68162.1"
/db_xref="GI:29604091"
/translation="MFAVXESRVRVGEIAATPTGALELVLHLSLGEVRESDDA
TARPDGLDFLELDPARGEWHTLALVRPVEGEEAGLFPFGAPESGSPARRICWRE
SAPRITDIEDPEVHRTLRSLDLGAPALHPAARPTHGGRRCPLVHGFPVRLAAV
LVHGTAAAGCTPR"
5183..5665
/note="SAV453"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68163.1"
/db_xref="GI:29604092"
/translation="WVTAVRARAGSAMQACRGARRSCHOSSHSEGVRRMDEKEFLRTV
AERTGLSRQEAADLTWATLETTLARLSPGEARDLTLPGLAEAVRRCTTDRIEFG

CDS
HHSVRRVAERTMLKEEADRGVTVLAVLREAISEKEFNLMSQLGTFDSQALASAG
5810..6964
/note="SAV454"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68164.1"
/db_xref="GI:29604093"
/translation="MIMTADVPETSSGSRVSRVALVADPGLPSEIARALAPLPDHLRR
AVGARQVQDMTAPLSQADVSDIAEAVRSHLDGAEDWDVAIFLTPDPRRLVYP
TSVEDTALRALISLALGVRELRPRVQAVDVVRELVAHEETPRHPVIGRPPE
PVDTRPFEQQRVWFGLGHTRLVAGMYANRPHRLFGLSLRALAGVPATATVLF
INSATNLATLALGAVCHAVITAGSAVALTMILIVDQWGERGTHLPARHFLYPNLV
TLVTIALAVVLAALFATVAVSFLLDASLRFKFGVQGVGDYFLFAWFTAIAM
IGGAFGTSLGDSKVSADAYGRQRQRMLRQMPRPERGERGPDITDG"
complement (6986..7498)
/note="SAV455"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68165.1"
/db_xref="GI:29604094"
/translation="MNPVAPASGIBQRPQSWLDEYLMWCSDDIADLAEGLEQRPLV
ATVAREAGERVLAVGPAQRVTRALGGRYMHQNAVAFGSPAFVGLGSMASMATSLG
NSARRAANDAPRCGARTPGLHVLVDVHGTMRLWSMHLGTSSRRPSSALRAL
ELVDHVTIAL"
7548..8189
/note="SAV456"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC68166.1"
/db_xref="GI:29604095"
/translation="WATLLRRVCPHEGVRIAAVPAAGFVLSRRPNTASVYRCVAVGPOS
EDRRVEGSRTPVASASRLRCPRRRLSVVRCPPRAVSPQVRSFRFRGTTRFLGVDV
VQRRDHALHARVIVIDVHPLGRLGARVRIATHSLGGVIAVDWATANVPLWTS
LVTFGSQATFFHPQVPGDLLCOMIAFRMRKIQEABTGALPPTGFIHAGQGV"
Query Match 91.6%; Score 17.4; DB 1; Length 299175;
Best Local Similarity 94.7%; Pred. No. 3.3e-03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCGACGCCCGCGATGACC 19
Db 187408 GTCGACGCCCGCGATCACC 187390
RESULT 14
AP005949 239850 bp DNA linear BCT 12-SEP-2003
LOCUS
DEFINITION
Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
15/31
ACCESSION
AP005949 BA000040
VERSION
AP005949.1 GI:27352050
KEYWORDS
Bradyrhizobium japonicum USDA 110
SOURCE
Bradyrhizobium japonicum USDA 110
ORGANISM
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE
1
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
Sasamoto, S., Watanabe, A., Iidesawa, K., Iriguchi, M., Wada, T., Yamada, M.,
Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)
JOURNAL
MEDLINE
22484998
PUBMED
12597275
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
Sasamoto, S., Watanabe, A., Iidesawa, K., Iriguchi, M., Wada, T., Yamada, M.
and Tabata, S.
TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)
JOURNAL
MEDLINE
22484998
PUBMED
12597275
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
Sasamoto, S., Watanabe, A., Iidesawa, K., Iriguchi, M., Wada, T., Yamada, M.
and Tabata, S.
TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)
JOURNAL
MEDLINE
22485002
PUBMED
12597279
AUTHORS
Kaneko, T.
TITLE
Direct Submission
SUBMITTED
120-NOV-2002 Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp)
URL: http://www.kazusa.or.jp/rhizobase/
Tel: 81-438-52-3935 (ex. 2398), Fax: 81-438-52-3934)
FEATURES
1..239850
Location/Qualifiers
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="Genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
18..1400
/genes="cysS"
18..1400
/genes="cysS"
/note="ORF ID:blr3790"
/codon_start=1
/transl_table=11
/product="cysteineyl-tRNA synthetase"
/protein_id="BAC49055.1"
/db_xref="GI:27352051"
/translation="MELRLYDLTKKRTFVFLDANNVVMYCGPTVYDFAHIGNARP
VIVFDVPLRLHLYGEAHVKVYRNITVDVDDKINDRAARDFPGLPLNEAIREVTQGT
KOPHADVPLGALRPSVEPRATEHIGEMREITERLVAGGFAYVAGDHVLFSPQANAA
NSTLPRYALSNSRLEMDVAGARVAVPKKSTDFVLWKPSKGPSPWSPSPAGTAAQ
GFPWIECSANAKWAKHGLYFDIHGGIDILVFPHENEVAQTCFCAHREMANVYMN
GFGVSEKSKSLGNFITHELLADWPGEVRLNKLKTHYRSPIDWTMKSLESAKT
LDWYRAADVDPGPAASVVEFLDDDLNTSLAALHGLRNSDVSLAGSLRLUGFL
SESQAQWEDRKQKASGVDAAGEVERLIISRTAARGRKDFKESDIRDQLAAMGVAIKDS
KEGTTWERSK"
1504..2007
/gene="blr3791"
1504..2007
/genes="blr3791"
/note="ORF ID:blr3791"
hypothetical protein
/codon_start=1
/transl_table=11
/protein_id="BAC49056.1"
/db_xref="GI:27352052"
/translation="MVKTWGQTLPRPGLRPFLDDVPVLAALIPVASIELTGEDYTEA
QQQAAVEAAESEEFGRKLAADLTLTATLESPVGFASLQGDYIRLYVHPVAVGGGI
ATMLVALBKLAGRGAKSLSDASDNAQGFPAKRGYVAMQNSITVNDENLMTMK
KTLGAPO"
2186..3799
/gene="blr3792"
2186..3799
/genes="blr3792"
/note="ORF ID:blr3792"
/codon_start=1
/transl_table=11
/product="2-isopropylmalate synthase"
/protein_id="BAC49057.1"
/db_xref="GI:27352053"
/translation="MSSAQAQKRLYVFTTLRDGAQNGVDTLTDKQIIAAMLDL
GIDYGVGGYGANPTDTEFGSKPKLNFARFTAFGNTRAGRSVNSNDPGVAGLLAKA
DAICFVAKSQAQVRVALETTEKENLASIRDSVAAKAAAGRMVMDCEHFFGQYKEDS
SPALACAAYAEAGRWVLCDTNGGDMPEVEAIVTEVTRI PGHVGIIHANNTEQ
AVANSLAAVRAGARQIQGTNLGLGRCGNALCSLPTLTKLQBEADAFEIGVTEKL
ATLVKYSRTLLDMLNEVNRHAAYGESAFTVTKGIHASAVLKDPQTYEHLVPELVGN
HRKVLSDQAGRSNVIAELDRAGIAYEKSDPKLTRVLEELKEREAGAYEASASFD
LLARRTLGKPHYFEVEQFDVNVQRYNALGERVTVALAVVKVDVAGEHLISNAEGNG

```

PVALDVALRKDLGKYQKIEGLKLDYRVRILNGGTGAVTRVLIESDENGDSTWTV
GVSNNIDASFOALMDSVIYKLVKGAPA"
3930. 4307
/gene="b1r3793"
3930. 4307
/gene="b1r3793"
/note="ORF ID: b1r3793
putative lyase"
/codon_start=1
/transl_table=11
/protein_id="BAC49058.1"
/db_xref="GI:27352054"
/translation="MIDHISVGVSDLDLRSATFYEAALGLRLVWRPRTVFGKAY
PEWNLREGMPVPVPSGVHICLRKATTAEDVAFHAAALSAGASDGPAGIRPHDRV
RYAAAFVDPDGNRIEAVTFPAE"
complement(4489. .5103)
/gene="b1r3794"
complement(4489. .5103)
/gene="b1r3794"
/note="ORF ID: b1r3794
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49059.1"
/db_xref="GI:27352055"
/translation="MAPMSTIKTVVCVCGSGPGTNPRFTGAKAFKALAEENNIRLVY
GGSLGMLGVSVTLVDNGGVTGIIPEFLRENNALTRVOEMIVPDMHERKRLMFE
RSDAFVALPGCVLEELVEQLTWKQGRHAKPVLANIINFWEPLFSLSHMRQTEF
IRAGLSVDILKADRVBEILPKLKAQAVALAEKELAPEVARKL"
5295. 7250
/gene="b1r3795"
5295. 7250
/gene="b1r3795"
/note="ORF ID: b1r3795"
/codon_start=1
/transl_table=11
/product="ABC transporter HlyB/MsbA family"
/protein_id="BAC49060.1"
/db_xref="GI:27352056"
/translation="MNQHPSPGADVPDPAGGPLELATLMTGLAHLWPHYWPQGRFDL
KRYVWGMVLLAAKLITLTPVFSKWATDALTGANTAPVLPDNLHLMVIAPELLTA
SYGATRLIMLQWRDLGFARVAMHVRKLAITTFHMHLSLREHLERKTKGLTRV
LEREGEIVTRVILQIPTIIVESLLVLLWQFDWRYVVALITVAVWYVTVI
ATEWYIGIRKMDSDTEANTKAIDSLINVEYKYFGAAREAOYDRSVARVEESSV
QAYSLAVNTGQAVITLGLTATMLMCAIGVNGTNGVDEVINAMMQLQPLNF
MGWYREIKQAIDIERMFGVIGRAEIKQDPQPLVISAGIVRFEDVRFATEPAPR
ILKGISFEVPAGTKVAIVGPGAGKSTISRLFLRYDISGRILIDQDIREVYQASL
RASIGMPQDTLVLFNDIRYNIRYRWDSADAEVEEARLAQIDNFIRMADPKYETQV
GERGLSGGKEQRVAITVILKAPPIVLDEATSDLTHTHEIIGALDRVAKNRTS
LVIAHRLSTIVGADEIIVLDQGRIAERGTAKLLAQGLVASMWRQREAEAREKLA
KWATSRAPNRQPPVEDALTPAAE"
7334. 8032
/gene="b1r3796"
7334. 8032
/gene="b1r3796"
/note="ORF ID: b1r3796
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49061.1"
/db_xref="GI:27352057"
/translation="MSLIDSTIQRIPTPHKEGYPIFGGFALASILFLWLSPLGWIGT
ILTVKCALFFDPDVRVTPVREGLVSPADGRVSMITWALPPAELGLGDRPLPRISVFM
SVFNCHVNRSPDIAGRDRIAYRPGLFINAEELKASEDNERNISLIVITPTARIGVQIA
GLIAKRVCFVREGCAIGAGERFGLIRFGSLDYLVPVGTKALVSEGTALAGETILA
DLADGDSRAYRAN"
8120. 9001
/gene="psaA"
8120. 9001
/gene="psaA"
/note="ORF ID: b1r3797"
/codon_start=1
/transl_table=11

```

```

/product="phosphatidylserine synthase"
/protein_id="BAC49062.1"
/db_xref="GI:27352058"
/translation="MTPYDIKDDPVRRRRERPIVVRMLVBNVITLAIACAGLTSLRLS
TEGRMSLAYAIVFAAALDGDIDRIARMIKQSGKGAELDSLADGVAGLWLYE
WQLHELGNAGWIAANVFASIGRLRNATMDNKKAPAFANFTGVPAPAGATVVL
LPIYVAFDLGRAPAAVTAATLLIAFLMWSRPLVFSKTKMRVPSLVLPAFVAVI
VFALIIAYPWHVLSIGTVLYLLALPLGYSYRQDQARDAATAAGBISSEPPSAPTLL
ANLSEPPHDDDRPGRSLH"
9175. 9942
/gene="b1r3798"
9175. 9942
/gene="b1r3798"
/note="ORF ID: b1r3798
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49063.1"
/db_xref="GI:27352059"
/translation="MPAASRQAAANLGENAVTDTATGPLPASVLEALGRYDPTTCNA
MEIVAPERRLIQYTTKQLVPPDPPIPVYGVARTVAINSVLKSSLPABEQSKRRIEY
EYVGTGHPRISVIQIDIDPDVGFAGVQSNVHKALGCLGVTIDGSIQIDQWAP
GFQALAGSIGPSHAWHAESEFGGEVVRVAGMTVKSDDLHADQHGAIVIPLDVAALKE
AAELCGRRETPILETARSPDFSLKLEKALKRSABIH"
10314. 11171
/gene="b1r3799"
10314. 11171
/gene="b1r3799"
/note="ORF ID: b1r3799
probable oxidoreductase"
/codon_start=1

```

Query Match 91.6%; Score 17.4; DB 1; Length 299850;
Best Local Similarity 94.7%; Pred. No. 3.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19
|||||
Db 250812 GTCGACGCCGCCGATGACC 250830
|||||

RESULT 15
AE017108
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) chromosome 10, section 62 of
77 of the complete sequence.
ACCESSION
AE017108 AE016959
VERSION
AE017108.1 GI:31432797
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriophytaceae; Oryzaceae; Oryza.
1 (bases 1 to 301666)

REFERENCE
AUTHORS
CONSRTM
TITLE
The Rice Chromosome 10 Sequencing Consortium
In-depth view of structure, activity, and evolution of rice
chromosome 10
Science 300, 1566-1569 (2003)
2 (bases 1 to 301666)
Buell, C.R., Wang, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
Direct Submission
Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
This is the pseudomolecule for rice chromosome 10, which was
constructed by resolving discrepancies between overlapping BACs,
trimming the overlap regions, and linking the unique sequences to
form a contiguous sequence. Genes in individual BAC clone were
identified by a combination of several methods: Gene prediction
programs, searches of the complete sequence against a peptide
database and EST databases. Genes with similarity to other proteins
are named after the database hits. Genes without significant

peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).

FEATURES
 source
 Location/Qualifiers
 1..301666
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="10"

misc_feature
 1..131983
 /note="Chromosome Sequence Derivation: nucleotide sequence in this region was derived from BAC clone OSUNBa0051D19 (GB:AC023240)."
 <2255..>2896
 /locus_tag="OSUNBa0051D19.24"
 /note="predicted by Genemark hmm"
 <2255..>2896
 /locus_tag="OSUNBa0051D19.24"
 <2255..>2896
 /locus_tag="OSUNBa0051D19.24"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP54385.1"
 /db_xref="GI:31432798"

gene
 /translation="NAGATMTGSRRLSHALLLVAILPNLAALAAVAGGGGGGFPF HSLTSSCQPSGALTITGDCNADGSECQGVQVMTYACPPVAAGGTGAALL HSLTADGAGSCPTGRFYDDQGLVALSTGFWDCRSCEKDWIRASGGASVTA MVVDCDORGCDSHNFEPCRNIVDGSAPVWDALGLNKDGEAQITWSDA"
 complement(<3441..>4021)
 /locus_tag="OSUNBa0051D19.23"
 /note="similar to metallothionein GI:3860191 (Silene vulgaris)"
 complement(join(<3441..3698,3833..>4021))
 /locus_tag="OSUNBa0051D19.23"
 complement(join(3441..3698,3833..4021))
 /locus_tag="OSUNBa0051D19.23"
 /codon_start=1
 /product="putative metallothionein"
 /protein_id="AAP54386.1"
 /db_xref="GI:31432799"

gene
 /translation="MAALEASGHWGFFRVGLRCGAPDCHRELFWVSGMGEVLA AVDLTVGVEKMAAFAGGEDGAGESPCTGDDGAEVAGDARDGRRRCRGGK HGROIHVSGVGVGEGAGESLLCTWELRERKRESGECEKVRPA"
 complement(<5481..>7359)
 /locus_tag="OSUNBa0051D19.17"
 /note="similar to Tam3 transposase GI:16063 (Antirrhinum majus)"
 complement(join(<5481..6777,6859..7041,7133..>7359))
 /locus_tag="OSUNBa0051D19.17"
 complement(join(5481..6777,6859..7041,7133..7359))
 /locus_tag="OSUNBa0051D19.17"
 /codon_start=1
 /product="putative Tam3 transposase"
 /protein_id="AAP54387.1"
 /db_xref="GI:31432800"

gene
 /translation="MGPTTLAQSSLRFNSVEGKGVSDTYTFDPAVARELLAAMIILH EPLGIQVHIGPRRVSALQPLFKVNTNTRKIDMTYEKKRALYAMAGASRVG LITDLWTNOKRGYMATATFIDSDTLRSIIMRFIVPAPTAEVICHLYEALVE WNLDSKISTLITDNTDVKLSLKIIGIKRKLMLGKLAHMECAHILNLIVDGL DVKDSIAKRESVATFATPKRKEFEIAKHVKVKNKGLDCKTRWNSTYRMLSD IALCARADRAIRKULFDCAPEEEWAFASEVVDRLKFNLDITAFVSGSYVTSNI QLLKTCEAKEQIKRWACGSDTIEMSVMTQKFDKWKWDIQGPMGLATLDPRFKID YLLGFIETIQGSEECATKVAEVDITLDMKEVEBDDDEDTESSAPLVDNLDLS SITAVTERRRPAIIVKVDLDELVSINTENFKILDMWKVAGTSFPTLRKVARD IFATIPVSTVASESASTSGRVLSEHRSRLTPELLEALMCSQDLRNKRYGTYDVLFIH HA"
 complement(<8562..>9337)
 /locus_tag="OSUNBa0051D19.16"
 /note="predicted by Genemark hmm"

mrna
 CDS
 gene
 mrna
 CDS
 gene
 mrna
 CDS
 gene

complement(join(<8562..8938,9148..>9337))
 /locus_tag="OSUNBa0051D19.16"
 complement(join(8562..8938,9148..9337))
 /locus_tag="OSUNBa0051D19.16"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP54388.1"
 /db_xref="GI:31432801"
 /translation="WAMGRTRSPVPPPPHRLPLDGDGNCVSMGNQWERKYPGSLGA GLAVRPPSPSPSVSEARYKSVRALAAPASAPAGRHARVAVRSSARPPQVP APRHAVFPFLPASQVRGRGLTGRTSASSTLRQYLIRHVGLFVLLMVDAGEGFG GKGGENDESTRRPAARVRHGFATTAKL"
 <11512..>14348
 /locus_tag="OSUNBa0051D19.22"
 /note="predicted by Genemark hmm"
 join(<11512..11580,11683..11742,11826..11894,12814..13188, 13280..13337,13752..13847,13959..14017,14109..14175, 14263..>14348)
 /locus_tag="OSUNBa0051D19.22"
 join(11512..11580,11683..11742,11826..11894,12814..13188, 13280..13337,13752..13847,13959..14017,14109..14175, 14263..>14348)
 /locus_tag="OSUNBa0051D19.22"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP54389.1"
 /db_xref="GI:31432802"

gene
 /translation="WAPAKRNLKAPIVDDSDSCDGLSRRLDIIIEIFRKEV VDKKEILNRKSCSHAFRYAKTVCDIFGLPSGLFEGKDFAGKEYILSMVGLSCLP SVRFQDQIKKEPLNEKVTITFLIACALCPNSLLIPSTKLTITFQDVNNUNRY DSKFTYDSMNMYMKFVNTSLGGLCNMAVYLDNVDFGNNYSNIPHQNVIEE NIENYGTSKADDDIGISAGISAOVHPVNDVLPSPHPIHNAFHGNEFAFVKSVEL NDSRQDDDLNFVTPQVGNANHSKQSDVLLDGLP"
 complement(<26683..>31718)
 /locus_tag="OSUNBa0051D19.15"
 /note="similar to gag-pol polyprotein GI:4234852 (Zea mays)"
 complement(join(<26683..30592,30833..31004,31574..>31718))
 /locus_tag="OSUNBa0051D19.15"
 complement(join(26683..30592,30833..31004,31574..31718))
 /locus_tag="OSUNBa0051D19.15"
 /codon_start=1
 /product="putative gag-pol polyprotein"
 /protein_id="AAP54390.1"
 /db_xref="GI:31432803"

gene
 /translation="MASARAVAMAMIFLLAALSTHLLASSLRPGAVAGGACRPSGYL PKGSYRAIMFLGSAARDYSLFVNITFYPTNNTGDKXKPTVNTNGNTASNSGG PLGSPSVVGFAAALPHADGNSRWKTRALLMLTAMQCFYVSRGKRSEPLSPPEE EAKFEASDCLFRGALISVLADNIDVYHMPSPGKMDRALEAKFGVSDGSELYMEQ FDYKAVDVSVEQAEHLOMLAKELNENCLPDKFVAGGIIAKLPSSWSPATSLK HRQFVSVDLIGSGVEEKAQDVGRKKVEGSSANVQKKNPHSNKKYKAPDV KPAATNFKKSGKAKGDFVYCGSGHWKDCPERKORKSANMIISBGGTSGYGI LFTLSVFSHPDWDTGANIHVACDILSPSSYQVGRGSSLLMGNSLAHVGVTV D LKFTSGKTVQLKNQVHPSPKKNLSVSGSLLCREDFLVPESNKKVYSGTGVGKYD SGLRPRFLSUNDNHNHNAVNHISENDSNVHSLCHVNFQGMTRLANMSLTPKFTLV KQSKCHTQVSKQPKPKHKAPEARNLAPLELHSDLCENMGVLTGKGGKFFMTLDDC TRFCYVYLLTKDEALHPFKYKAEVQJLERKIKELSDRGGVYFSEFAPCFCEFG IHEMTPTPSPQSGVAERKKNTIEMVNMALDTAGLSKEMWGEAVLTACHVNLIPM KKEVTFPFEWERKLNLSYLTWGLCAKVNVIKAKRKLGPKTVCVFLGTAIHSVD YRFLVNSGPDHVTGTTLESRDATFENEFPMKYTPSTSKETVMPHEFAPIBHND QPPEENEDNTVTRKSRQVAKSGFDYTVLVDPTPTTIEEYSPADYKWEA QVSEMDISNGTWEVPRPYGKPIGCKVFKKLPDGTIEKYKALVAKYQKE GDFDTSPYVARLTITRVLAAASHGLVHQMVKTAFLNGELEEBEIIYMDQPGYV LBGQEGMCKLLSLYGLKQAPKHEKPDFTLTSGAFVNAEDCVVYVGGGSEVI LCQVDDIIFGTSNIEVDKYLKSKFEMKDLSEADVILNKLQREGGKITLQVS HYVDKVLSPGFSYDNDHWRALERMYRLQYSGIYHYTGYPKVLGYSNWSID ISFAGSKSRFSVNFDDHWRALERMYRLQYSGIYHYTGYPKVLGYSNWSID ADEIKATSGYVFTLGGAVSWKSCQTLITRSTMEALTDATVAEWRELLMDL PVYERKVPAILANCONQIVIIKVNSSKNMKSRSHIKRRLKSVRKQKNSGVIALDYO TARNLADQFTKGLPCNVIDASREMLIPT"
 complement(<32357..>35641)
 /locus_tag="OSUNBa0051D19.14"
 /note="predicted by Genemark hmm"

```
mRNA      complement(join(<32357. 32362,32539. 32932,33018. 33121,
35012. 35641))
/locus_tag="OSUNBa0051D19.14"
CDS        complement(join(32357. 32362,32539. 32932,33018. 33121,
35012. 35641))
/locus_tag="OSUNBa0051D19.14"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP54391.1"
/db_xref="GI:31432804"
/translation="MAANVKVLVVLALLQLMSLHVVHGDNGGVSAAVATGKHEPKPK
QTGGGGGGGCHISGLHGKAGKCNRAHSGDCVTRRYPQRCPPVSVAPPTPAT
LTSNFRAGGDCGRSSCDGRFHPDPTNVALSSGWLRLDGCRCNMRIRVAAGPAT
ALAEVDECDVNSGDCGAEHNFEPCCPNVDVDSGPAVKALDGEQGEKFAKSLSTS
PTIHVSDDAASCHASGVYVHLGNETFTTALITLKVFDHGEDDGGVPTSCDMRFERN
TELVAUSSWLRGGRRCHRRIRTFVAVAGASGRSSVAVRVDDCDVSNGCREED
GFAPPCRNNAVGSPPVWEKLGNASVEFEVMSCLKNAKEL"
gene        complement(<39112. >39813)
/locus_tag="OSUNBa0051D19.13"
/notes="similar to secretory protein GI:5669008(Triticum
aestivum)"
mRNA        complement(<39112. >39813)
CDS          complement(<39112. >39813)

Query Match      91.6%; Score 17.4; DB 8; Length 301666;
Best Local Similarity 94.7%; Pred. No. 3.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTCAGCGCGCGGATGACC 19
|||||
Db 193940 GTCAGCGCGCGGACGACC 193958

RESULT 16
BX640451 343473 bp DNA linear BCT 14-AUG-2003
LOCUS      Bordetella bronchiseptica strain RB50, complete genome; segment
DEFINITION 15/16.
ACCESSION  BX640451 BX470250
VERSION     BX640451.1 GI:33577672
KEYWORDS    complete genome
SOURCE      Bordetella bronchiseptica RB50
ORGANISM    Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Alcaligenaceae; Bordetella.
REFERENCE   1 Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N.,
            Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
            Mungall,K.L., Cerdeno-Farraga,A.M., Temple,L., James,K., Harris,B.,
            Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bason,N.,
            Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
            Doggett,J., Feltwell,T., Goble,A., Hamlin,N., Hauser,H.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Norberczak,H.,
            O'Neill,S., Ormond,D., Price,C.S., Rabinowitsch,E., Rutter,S.,
            Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
            Skelton,J., Squares,R., Squares,S., Stevens,K., Unwin,L.,
            Whitehead,S., Barrall,B.G. and Maskell,D.J.
            Comparative analysis of the genome sequences of Bordetella
            pertussis, Bordetella parapertussis and Bordetella bronchiseptica
            Online Publication
JOURNAL     Nat. Genet. DOI:10.1038/Ngl227-10.1038/Ngl227
REMARK      2 (bases 1 to 343473)
AUTHORS     Sebahia,M.
TITLE       Direct Submission
JOURNAL     Submitted (06-AUG-2003) Submitted on behalf of the Pathogen
            Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, E-mail: ms@sanger.ac.uk
FEATURES    Location/Qualifiers
            1..343473
             /organism="Bordetella bronchiseptica RB50"
             /mol_type="genomic DNA"
             /strain="RB50"

/db_xref="taxon:257310"
290. 1084
/locus_tag="BB4576"
290. 1084
/locus_tag="BB4576"
/notes="ortholog of Bordetella pertussis (BX470248) BP3319"
/codon_start=1
/transl_table=11
/product="putative IclR-family transcriptional regulator"
/protein_id="CAE34938.1"
/db_xref="GI:33577673"
/translation="MRCGIRHMDXLLKGLTVLFAIRSLDQATFTLDQIAEEICGLTR
SNTHRTLOTHAGVYRDRDPATGSYQCTIKLFELGARQLASIDARKLAPPFNGEILARL
NETVHSLVDGLDVVYDKLDSPOPLLAYSIVGGRAYAVATGKALLACOTADYLD
HYGAALKHETDTHVLLAKDELADNVRIGVAINRGWROGIGIAAIFNGLDKAV
AALGISGFLERLTNQKREYARLVQQAARQLSAHNGYRGYFGGA"
misc_feature 494. 1045
/locus_tag="BB4576"
/notes="HMMpfam hit to PF01614, DE Bacterial
transcriptional regulator, score 1.3e-54"
1101. 2261
/locus_tag="BB4577"
1101. 2261
/locus_tag="BB4577"
/notes="ortholog of Bordetella pertussis (BX470248) BP3318"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAE34939.1"
/db_xref="GI:33577674"
/translation="MNQESTLAGVRVIEICNVAGPFCOMLLADMGADVVKLEPPGTG
DTRSWAPITDGYSENFLNKRKSVTLNKNRABERDAIIEIAGADVLIENRPGV
MDRLGLGQGAARNRPLVYCSVAYSGSPRSGDFDLTIQANGLSMVTSVEEGA
PVKCGVPVCDFTAGLYAAFAI AALRRVQADQCGTHIDVSMGLGATLGAALQTSYFG
TGADPVKLGSAPHRNAPQVFRCKNGYFGMAAGNLSMKAVCKVVERDDLHADPRFAN
PDRAHOVALRDILEAFEAHDSHAWLERPQAGVPQCAPINDYSEVLADPOVHMDW
VDLPLNGVTRTFVSPVPRTAGVWRPFPALGEHNDELISALAARPAQ"
1311. 1877
/locus_tag="BB4577"
/notes="HMMpfam hit to PF02515, DE CAIB/BAIF family, score
6.1e-59"
2258. 3001
/locus_tag="BB4578"
2258. 3001
/locus_tag="BB4578"
/notes="ortholog of Bordetella pertussis (BX470248) BP3317"
/codon_start=1
/transl_table=11
/product="putative enoyl-CoA hydratase/isomerase"
/protein_id="CAE34940.1"
/db_xref="GI:33577675"
/translation="MSAMSASLPLAIERRPAAWTTLSRPEKRNALSALVELALIDGV
DAAREQVPLVFAAGRNFSAGFDFTDYETSEGDLRLRMVRIEMLLQRVAGSPSLT
LALAHGRNFGAGVDLFAACKWRYCTPEAGFRMFLGLVGLTRFRFDIVGADQALSI
LGSARFDADQARRIGVRDCAQAQWPAIDAAAATAALDPATRLHRLVRDDHD
DADLAALABSAAPGLKARIDYLAQPA"
2312. 2806
/locus_tag="BB4578"
/notes="HMMpfam hit to PF00378, DE Enoyl-CoA
hydratase/isomerase family, score 7.7e-10"
complement(2998. 3873)
/genes="copB"
/locus_tag="BB4579"
complement(2998. 3873)
/genes="copB"
/locus_tag="BB4579"
/notes="ortholog of Bordetella pertussis (BX470248) BP3316"
/codon_start=1
/transl_table=11
/product="copper resistance protein"
/protein_id="CAE34941.1"
/db_xref="GI:33577676"
/translation="NAASUGLAAAAASAAQAQAPAMDHGSMGMDHGAAPASAMDHGSM
```



```

/note="ortholog of Bordetella pertussis (BX470248) BP3166"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39135.1"
/db_xref="GI:33574490"
/translation="MHDISSLTTRKTLPLRVLGVGVGPGSGKTTLLVMCKWAMYPOF
DLIAITNDIVYKEDQRLTLTSGALPPERILGVETGGCPHTAIRDASINLIADQMLE
QPADIVFVYSGDNLAAATFSPSLDLTIYIDVASEGKIPKGGPGITKSLDFIN
KTLAPYVGADLAVMEADTRMRGDKPFCVNLKTGDLQDVIAFLKTEGLFRG"
complement(509..892)
/misc_feature
/feature="ureg"
/locus_tag="BPP3852"
/note="HMPfam hit to PF01495, HypB/UreG
nucleotide-binding domain, score 4.5e-73"
complement(848..871)
/misc_feature
/feature="ureg"
/locus_tag="BPP3852"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."
complement(962..1624)
/feature="uref"
/locus_tag="BPP3853"
complement(962..1624)
/feature="uref"
/locus_tag="BPP3853"
/note="Ortholog of Bordetella pertussis (BX470248) BP3167"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39136.1"
/db_xref="GI:33574491"
/translation="WHLSSPALPIGFGFSQGLEAAELGLVHDEASTLAWIPSQLVLT
VMAEAPLACLLFEANRAGDDAAAGHWNQFHASRETRRLROETEQMGSLARLAQE
LWGTAALRAALRAATLPAFHACAMWALPREAGLGAYFVWLENOVAAAIKGV
PLGQWAGRMLEURLAGLPAVLADARARAGATPRLDTFAPQVALVSARHETOPSLRF
RS"
complement(1094..1546)
/misc_feature
/feature="urep"
/locus_tag="BPP3853"
/note="HMPfam hit to PF01730, UreF, score 9.7e-74"
complement(1647..2264)
/feature="ureb"
/locus_tag="BPP3854"
complement(1647..2264)
/feature="urep"
/locus_tag="BPP3854"
/note="Ortholog of Bordetella pertussis (BX470248)
BP3167A"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39137.1"
/db_xref="GI:33574492"
/translation="MKIANFIKGAAGGLDLSQAPGVTLAERRRSRORLDSEGR
GELGMALRGTLRDGDLVVAEDGTYYVVRALDEDVAVTAATPWQLARAAHILGNRH
VLLIAEQHLOFEYDVLIDMLAQLGVTAMRLRAVPEPDVYAGGGHGHDESFQD
DYALAAQYHAHAEPHAPHAGHGHHVHSGHGHGKHGHEAAS"
complement(1944..2249)
/feature="ureb"
/locus_tag="BPP3854"
/note="HMPfam hit to PF02814, UreE urease accessory
protein, score 1.7e-45"
complement(2267..3982)
/feature="urec"
/locus_tag="BPP3855"
complement(2267..3982)
/feature="urec"
/locus_tag="BPP3855"
/EC_number="3.5.1.5"
/note="Ortholog of Bordetella pertussis (BX470248) BP3168"
/codon_start=1
/transl_table=11
/product="urease alpha subunit"

```

```

/protein_id="CAE39138.1"
/db_xref="GI:33574493"
/translation="NTRISRSAYAIYGVTVVGGVDRVRLADTILLAEVEKDHITFG
EVKFGGKVIKRDGMQORLATCDVTVITNALLIIDA VGTIVKADIGIKGLGISGIG
KAGNPTQGVITIIIGASTEVVAGGLIVTAGAIDTHIFICPOQIEBALATGTTMI
GGGTGPATSLATTSTSPFWMAAMQLADAPFVNVVGLFGKSSSHGALLLEQVRAGA
MGLKIHEDWASTPASIDICLVNAEETDIOVAIHSDTLNESGFEVDITFAFKGRTIHSF
HTEGAGGHAPDIIRAAGPNVLPASTNPTWFTNTIDHLDMMVMVCHLDPISAED
LAFASRRIRRTIAEEDILHDGAFSIMSSDQAMGRVGEIVLRTQTAHKNKLQKGP
LOQDSRNSERIKRYIAKTYINPAVAGIAHLGVSEVGLADLVLPKPAFFGVKVN
MYLKSGMAVSASIGDMGASISTPQVQIRPMKSHGKALRTSVAFVSOVSLSPAVSE
LELNKLEAVRCRGVTKHDMYRNWNLPAISVDPTQYVYADGQLLRCEALAELEPMAQ
RYLFLP"
complement(2273..3580)
/misc_feature
/feature="urec"
/locus_tag="BPP3855"
/note="HMPfam hit to PF02802, Urease alpha-subunit,
catalytic domain, score 0"
complement(2969..3019)
/misc_feature
/feature="urec"
/locus_tag="BPP3855"
/note="PS00145 Urease active site."
complement(3548..3589)
/feature="urec"
/locus_tag="BPP3855"
/note="PS01120 Urease nickel ligands signature."
complement(3584..3979)
/misc_feature
/feature="urec"
/locus_tag="BPP3855"
/note="HMPfam hit to PF00449, Urease alpha-subunit,
N-terminal domain, score 4.8e-83"
complement(3979..4287)
/feature="ureb"
/locus_tag="BPP3856"
complement(3979..4287)
/feature="ureb"
/locus_tag="BPP3856"
/EC_number="3.5.1.5"
/note="ortholog of Bordetella pertussis (BX470248) BP3169"
/codon_start=1
/transl_table=11
/product="urease beta subunit"
/protein_id="CAE39139.1"
/db_xref="GI:33574494"
/translation="MIPQELTPEGQIELNVGRPTLTIAVNEGRPIQVGSYHFAE
ANNAVDFRELATGYRLNIPAGNAVRFPEGMERTVELVAVGGERRIFGQGVMGALK"
complement(3985..4284)
/misc_feature
/feature="ureb"
/locus_tag="BPP3856"
/note="This hit extended beyond the end of the feature by
1 aa and was clipped.
HMPfam hit to PF00699, Urease beta subunit, score
3.2e-60"
complement(4304..4921)
/feature="urej"
/locus_tag="BPP3857"
complement(4304..4921)
/feature="urej"
/locus_tag="BPP3857"
/note="ortholog of Bordetella pertussis (BX470248) BP3170"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39140.1"
/db_xref="GI:33574495"
/translation="MKAMSKRAMLGSGRAALMLFSGAALAHPHGLGHELPGSMFAAGF
WHPLVGFHLLANLAVGHWMSALHTHTARCAVWLPVWFALLPAGMMGAGVRLPAVE
PVMVSLVILGLLVASRKAVQWAGFALGVGFALPHGLAHGMEIPLGSEGALGFVAGFM
LATLGLHLAGLPAFGLKFKLKNLWLSRALGVGIAGYCALLFVGARY"
complement(4844..4921)
/misc_feature
/feature="urej"
/locus_tag="BPP3857"

```

```

/note="Signal peptide predicted for BPP3857 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 1.000 between residues 26 and 27;
signal-peptide site"
complement(join(4310..4375,4397..4462,4493..4549,
4571..4624,4637..4702,4724..4789,4835..4900))
/genes="ured"
/locus tag="BPP3857"
/note="7 probable transmembrane helices predicted for
BPP3857 by TMHMM2.0 at aa 7-29, 44-66, 73-95, 100-117,
91.6%; Score 17.4; DB 1; Length 346259;
Best Local Similarity 94.7%; Pred. No. 3.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

misc_feature
1 GTCGACGCCGCGGATGACC 19
|||||
Db 333562 GCCGACGCCGCGGATGACC 333580

RESULT 18
BX640412/c
LOCUS
DEFINITION
BX640412 BX470248
VERSION
BX640412.1 GI:33571216
KEYWORDS
complete genome.
SOURCE
Bordetella pertussis Tohama I
ORGANISM
Bordetella pertussis Tohama I
Bacteria: Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
REFERENCE
1 Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N.,
Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
Mungall,K.L., Cerdeno-Tarraga,A.M., Temple,L., James,K., Harris,B.,
Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bason,N.,
Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
Doggett,J., Feltwell,T., Goble,A., Hamlin,N., Hauser,H.,
Holroyd,S., Jagers,K., Leather,S., Moule,S., Norberczak,H.,
O'Neill,S., Ormond,D., Price,C., Rabinowitch,B., Rutter,S.,
Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
Skilton,J., Squares,R., Squares,S., Stevens,K., Unwin,L.,
Whitehead,S., Barrrell,B.G. and Maskell,D.J.
Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
Online Publication
Nat. Genet. DOI:10.1038/Ngl227-10.1038/Ngl227
2 (bases 1 to 348171)
Sebahia,M.
Direct Submission
Submitted (06-AUG-2003) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: mss@sanger.ac.uk
Location/Qualifiers
1..348171
/organism="Bordetella pertussis Tohama I"
/mol_type="genomic DNA"
/db_xref="taxon:25713"
complement(52..921)
/genes="bhuT"
/locus tag="BP0345"
complement(52..921)
/genes="bhuT"
/locus tag="BP0345"
/note="Similar to Yersinia enterocolitica hemin binding
protein HmT TR:P74979 (EMBL:X77867) (279 aa) fasta
scores: E(): 8.2e-29, 37.638% id in 271 aa, and to
Yersinia pestis hemin-binding periplasmic protein HmUT
precursor SW:HMUT_YERPE (Q56591) (279 aa) fasta scores:
E(): 9.4e-29, 37.638% id in 271 aa"
/codon_start=1

TITLE
JOURNAL
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
/note="Previously sequenced as Bordetella pertussis outer
membrane heme receptor BhuR TR:AAK38153 (EMBL:AY032627)
(149 aa) fasta scores: E(): 2.7e-49, 100.000% id in 149
aa. Similar to Rhizobium meliloti putative iron transport
protein Smc02726 TR:CAC46967 (EMBL:AL591790) (743 aa)
fasta scores: E(): 4.6e-37, 37.630% id in 768 aa"
/codon_start=1
/transl_table=11
/product="outer membrane heme receptor"
/protein_id="CAE44679.1"
/db_xref="GI:33571219"
/transl_table=11
/note="Similar to Yersinia enterocolitica hemin binding
protein HmT TR:P74979 (EMBL:X77867) (279 aa) fasta
scores: E(): 8.2e-29, 37.638% id in 271 aa, and to
Yersinia pestis hemin-binding periplasmic protein HmUT
precursor SW:HMUT_YERPE (Q56591) (279 aa) fasta scores:
E(): 9.4e-29, 37.638% id in 271 aa"
/codon_start=1
/transl_table=11
/product="putative hemin binding protein"
/protein_id="CAE44677.1"
/db_xref="GI:33571217"
/transl_table=11
/note="Signal peptide predicted for BP0345 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.617 between residues 23 and 24;
signal-peptide site"
complement(148..840)
/genes="bhuT"
/locus tag="BP0345"
/note="HMPfam hit to PF01497, Periplasmic binding
protein"
complement(835..900)
/genes="bhuT"
/locus tag="BP0345"
/note="1 probable transmembrane helix predicted for BP0345
by TMHMM2.0 at aa 7-23"
complement(918..1964)
/genes="bhuS"
/locus tag="BP0346"
complement(918..1964)
/genes="bhuS"
/locus tag="BP0346"
/note="Similar to Yersinia pestis hemin transport protein
HmUS SW:HMUS_YERPE (Q56590) (345 aa) fasta scores: E():
1.1e-46, 42.486% id in 346 aa, and to Yersinia
enterocolitica hemin transport protein HmS SW:HMS_YEREN
(Q31517) (345 aa) fasta scores: E(): 1.2e-46, 43.413% id
in 334 aa"
/codon_start=1
/transl_table=11
/product="putative hemin transport protein"
/protein_id="CAE44678.1"
/db_xref="GI:33571218"
/transl_table=11
/note="Previously sequenced as Bordetella pertussis outer
membrane heme receptor BhuR TR:AAK38153 (EMBL:AY032627)
(149 aa) fasta scores: E(): 2.7e-49, 100.000% id in 149
aa. Similar to Rhizobium meliloti putative iron transport
protein Smc02726 TR:CAC46967 (EMBL:AL591790) (743 aa)
fasta scores: E(): 4.6e-37, 37.630% id in 768 aa"
/codon_start=1
/transl_table=11
/product="outer membrane heme receptor"
/protein_id="CAE44679.1"
/db_xref="GI:33571219"
/transl_table=11
/note="Similar to Yersinia enterocolitica hemin binding
protein HmT TR:P74979 (EMBL:X77867) (279 aa) fasta
scores: E(): 8.2e-29, 37.638% id in 271 aa, and to
Yersinia pestis hemin-binding periplasmic protein HmUT
precursor SW:HMUT_YERPE (Q56591) (279 aa) fasta scores:
E(): 9.4e-29, 37.638% id in 271 aa"
/codon_start=1

```



```

misc_feature      YEQCILLAP"
                  complement(5734..5799)
                  /gene="huri"

Query Match          91.6%; Score 17.4; DB 1; Length 348171;
Best Local Similarity 94.7%; Pred. No. 3.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1   GTGCAGCGCCGCGATGACC 19
         |||||
Db       15873 GCCGACGCCGCGATGACC 15855

RESULT 19
ADP279815/c
LOCUS      ADP279815
DEFINITION Ascovirus DpAV4 Peti genomic DNA fragment, ORF1-ORF15.
ACCESSION AJ279815
VERSION    AJ279815.2 GI:14031079
KEYWORDS   ORF1; ORF10; ORF11; ORF12; ORF13; ORF14; orf15; ORF2; ORF3; ORF4;
            ORF5; ORF6; ORF7; ORF8; ORF9.
SOURCE     Ascovirus DpAV4
ORGANISM   Ascovirus DpAV4
VIRUSES; dsDNA viruses, no RNA stage; Ascoviridae; Ascovirus.

REFERENCE 1
AUTHORS    Stasiak,K., Denattei,M.V., Federici,B.A. and Bigot,Y.
TITLE      Phylogenetic position of the Diadromus pulchellus ascovirus DNA polymerase among viruses with large double-stranded DNA genomes
JOURNAL    J. Gen. Virol. 81 (Pt 12), 3059-3072 (2000)
MEDLINE    20540044
PUBMED     11086137
REFERENCE 2
AUTHORS    Bigot,Y.
TITLE      Direct Submission
SUBMITTED  (19-JAN-2000) Bigot Y., IRBI, Groupe d'etude des
Parasites moléculaire, faculté des sciences, Parc Grandmont, Avenue
Monge, 37200 Tours, FRANCE
REVISED BY [4]
3 (bases 1 to 6736)
Bigot,Y.
Direct Submission
SUBMITTED  (08-MAY-2001) Bigot Y., IRBI, Groupe d'etude des
Parasites moléculaire, faculté des sciences, Parc Grandmont, Avenue
Monge, 37200 Tours, FRANCE
ON MAY 14, 2001 this sequence version replaced gi:11931735.
RELATED SEQUENCES: ORF3 - Q00196 (SP), ORF2 P14358 (SP), ORF1 -
O55720 (SP), ORF5 - O55733 (SP), ORF6 - O55763 (SP), ORF7 -
CEF19B2_2.

FEATURES             Location/Qualifiers
     source           1..6736
                     /organism="Ascovirus DpAV4"
                     /virion
                     /mol_type="genomic DNA"
                     /specific_host="Diadromus pulchellus"
                     /db_xref="taxon:113365"
                     /lab_host="Diadromus pulchellus"
                     complement(<1..990)
                     /note="ORF2"
                     Fv3 iridoviral protein homologue"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAC19145.1"
                     /db_xref="GI:11931737"
                     /db_xref="SP:REBEL_Q09S0U"
                     /translation="VAIITKSLAIERTTEDAADHGIRITDELGLNFCDAMPPTDLD
ERKRVGVDPDGKLVRFSPPFTPLGPESLDGKFVTMKEGTVIRVFPHNTKWVSFN
RKIDAFKXWGIAISFEIEFTNFKLLGSKDEFLSDKSICYLFMRGTIPSTRIVA
PAYDLIELNSFDMSNGSGSKIENLSLSYNTKOFNSLNKDVEEYNWMPFPDSGCVF
FLYNPDACEKLENSYTKLSELRHNVSSTPFAYLHNCFDENNKKFRDLYBEHVETFD
FDSEINALSANLLNYIIKHVDQKIEVSKGEHVLVYHHVGKLYSSRPQTTLQ"
                     560..1039
                     /note="ORF3"
                     murine protein homologue"

CDS           1..6736
              /note="ORF3"

```

```

/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19146.1"
/db_xref="GI:11931738"
/db_xref="SPTREMBL:Q9DST9"
/translation="MLSSDSKNSFDFPNSFVTLKNSPKEMAPHLDLKASSFLLE
LTVHLVAKKTLITVPSFMVTKLPESRLSGSPGVGKGDGLNTNLSPSNTTPTRIUSRS
VVGMSAQKFRSPSSVLLPWSAASSVRSINALFVILAMFWVNLITSEFFNRP"
complement(1722..2318)
/notes="ORF4"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19147.1"
/db_xref="GI:11931739"
/db_xref="SPTREMBL:Q9DST8"
/translation="MINKVSRPCALDADPSIKTSVCLRSASVLLSWHLNITVEES
WSMSRSLTUPPVELTKANVVFVFLHSFMDRPFVITRLIMLWLTLLVLAASCAP
SPSTRMADCSFSPFSLTLLVELSAFNKLP/LTLTRCKVTLFMLTSPVPFSIFTM
DWLLSEYAFICFVLATSVTAFTILVELVDIADPIYII"
1735..3060
/notes="ORF5"
CIV iridoviral protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19148.1"
/db_xref="GI:11931740"
/db_xref="SPTREMBL:Q9DST7"
/translation="WGAISTNSTEIVMKAVTEVANTNIOKAYSDNNQSIIVKIENTG
GDVNNITFORVNVNNGVNLNLTNTNEVNEKLQKASMLVEGLNLAQLAAS
NVNVLINLVTKNRSINECKNSTKTFSDVRNTGGVNVNRDLIDQSDSVFKCI
QDSKTLADRLQTDVLDGLASSKAQGLDRLVIGVGFALGVOGTFIGAKVIGP
VIASPELLASTDNTNSPRTQIKTNKVKSLKNATPTATFTPAQILSKYKSAEN
LTDIILGADVAVLEGTVLYKSAKPRDAYDAPDDGLGPVLEIIIDPVNTRGI
SVKLNRGLANFANPVLNLIKREDMAAGDVAGVQDOSTHTVSVFKEDDPTFTEN
VPASVMIKPEKPGVKNQLVYIGYALLAIGMATINSFRNGSGGTRS"
3171..3569
/notes="ORF6"
CIV iridoviral protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19149.1"
/db_xref="GI:11931741"
/db_xref="SPTREMBL:Q9DST6"
/translation="WMPYSAPKLTDELRIIGDKEVQIVIVEROMVQSHAIYRVASHT
ENYFAMVSAEYIVCDPKLDGLGVGKYAARKTKABCAKKYQEGMLDLASIDA
HKKDDMDCAICQWLWFCANRVKFNVTIK"
4004..4315
/notes="ORF7"
nematod protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19150.1"
/db_xref="GI:11931742"
/db_xref="SPTREMBL:Q9DST5"
/translation="MLKNVTTCIFRALSSFWILSIWISHFLLMWICLAAEMGMLO
LTTSERFSLKLYLLSMGLTMSLTIVIPGSLNMGDPVFRNCKSVLFGVDNALIL
V"
complement(4301..4664)
/notes="ORF14"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC38378.1"
/db_xref="GI:14031081"
/db_xref="SPTREMBL:Q9DST4"
/translation="MRRPRYFRSNTFOINNFRGEVNNVVARLIHETTFENFVSIL
KSKMLYKPSNNLNGTVQGSRRNRIVSDPFIISLKNKNFYDYDEVDAVMRLYDEDTF
VHDDVMFIFDPDSSIPV"
complement(4302..4664)
/notes="ORF8"
ORF9"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19151.1"

```

```

/db_xref="GI:11931743"
/db_xref="SPTREMBL:Q9DST4"
/translation="MRRPRYFRSNTFOINNFRGEVNNVVARLIHETTFENFVSIL
KSKMLYKPSNNLNGTVQGSRRNRIVSDPFIISLKNKNFYDYDEVDAVMRLYDEDTF
VHDDVMFIFDPDSSIPV"
complement(4670..5072)
/notes="ORF14"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC38379.1"
/db_xref="GI:14031082"
/db_xref="SPTREMBL:Q9DST3"
/translation="MEWKIDAWVFNCRVLLFREFTDAFKNASVDSIRLVQKTL
QRTVDLEAPITDDVNMMLFETPAATVKLVEALLFEATQALDSKRRQKIYSVVGVED
LJRDVMSLTSEDKTPSORHYEGDRFNQV"
complement(4671..5072)
/notes="ORF10"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19152.1"
/db_xref="GI:11931744"
/db_xref="SPTREMBL:Q9DST3"
/translation="MEWKIDAWVFNCRVLLFREFTDAFKNASVDSIRLVQKTL
QRTVDLEAPITDDVNMMLFETPAATVKLVEALLFEATQALDSKRRQKIYSVVGVED
LJRDVMSLTSEDKTPSORHYEGDRFNQV"
4888..5157
/notes="ORF11"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19153.1"
/db_xref="GI:11931745"
/db_xref="SPTREMBL:Q9DST2"
/translation="MFQSTVICASRSTSVRCRVFWTRTLMESTDAFFLKASVNSLN
NTRLQNFHASIFHSMKNPSETLDSRSRAVTPRSLMGRTHL"
complement(5437..6047)
/notes="ORF1"
CIV iridoviral protein homologue"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAC19144.1"
/db_xref="GI:11931736"
/db_xref="SPTREMBL:Q9DSU1"
/translation="AGRTSTRGAGRGORPHELFDRLGILQIORRSOLYARKTETR
DHSILVGVLSLQPTNHLHEHKGQGGRRRESVLEGRHQNGKRPYGFPIPGQ
FNTSQIYDASGGSGFGRISTGRSRRARRARFVNDLHTSNVVKCKDKNLFIL
YFPDFAKRVKLVTRGYPIYIDYLSYSESCDGMASRDL"
complement(join(5440..6079,6404..6513))
/notes="unnamed protein product; ORF15"
/codon_start=1
/protein_id="CAC38841.1"
/db_xref="GI:14132804"
/db_xref="SPTREMBL:Q9IU95"
/translation="MFSSAIHPALRPQNCULLIAEGLTVLLQLSTLSKRSTGFVR
TLEATGRTSRRGAGRGORPHELFDRLGILQIORRSOLYARKTTRDSDSLVGLV
SLSQPOTNHLHEHKGQGGRRRESVLEGRHQNGKRPYGFPIPGQFNTSQIYD
ASGGSGFGRISTGRSRRARRARFVNDLHTSNVVKCKDKNLFILYRFDFAKGR
VALVTRGYPIYIDYLSYSESCDGMASRDL"
complement(5440..6079)
/notes="ORF15"
/number=2
complement(5523..5849)
/notes="ORF12"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19154.1"
/db_xref="GI:11931746"
/db_xref="SPTREMBL:Q9DST1"
/translation="MKNIKVAKADGASPKVATKTVNVDLIGFIFYLDNFTLHKF
IRTHPAEEVHSVVFQVIALVVRAGALSCITCTLPTWLSKNVIKISYCTASTSRV

```

Query Match 89.5%; Score 17; DB 14; Length 6736;
 Best Local Similarity 100.0%; Fred. No. 6.3e+03; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 0;

[illegible][illegible]


```

/ gene="OJ1065_B06.1"
/ note="supported by full-length cDNA(s): AK105640"
1766..3334
/ gene="OJ1065_B06.1"
/ note="contains full-length cDNA(s): AK105640"
similar to Arabidopsis thaliana chromosome 2, NP_850414"
/ codon_start=1
/ product="unknown protein"
/ protein_id="BAC83058.1"
/ db_xref="GI:34393497"
/ translation="MLTGLSFRYSVDGGFAKGSIDLGLGVROVTTFAKWTSGDGG
CATPRPQVAGSALGHYAQRNDRLPGHVLVAROVSGGLIAPLPLDVPYVSSOD
DAHFWLTPDGGRAIGAVATSPDKPREVEACVCRADFTDACEAATVWDKDGPSA
LALPAVGVARGVHAGTFFVLARSATAASALACLKNNAAATSCMLPQAGNAL
LAARPAQLFHLDEPFLPSSVTFWFGALLYKQSGTPTPEAADGNSLPGQGGNDGG
YLDLPVDFORERVKKGLAGAKVYQAKEMLGATATDLAVWFYFNFNGPARKVGP
LTPLGKTGGEHWEHVTLRVSNFSGVQLKRWFSQISAGAWDAPOLEYLDGNRPS
AVSSIHGALYPRAGLVQGDARLGVGRNDCDRGLDTGGACRCVSAEYLGGGG
GGVAREPTWLLSDRENGPREEDYDGREINREKVLPRSTRERLKLVSFVGEQPTGP
RMKGSWRNDEAREAK"
4131..11355
/ gene="OJ1065_B06.2"
/ join(4131..4486,5711..5834,5914..6058,6190..6415,
7021..7094,7268..7360,8691..8808,9598..9660,10032..10064,
10583..10659,10807..11355)
/ gene="OJ1065_B06.2"
/ note="supported by full-length cDNA(s): AK072012"
/ join(4248..4486,5711..5834,5912..6058,6190..6415,
7021..7094,7268..7360,8691..8808,9598..9660,10032..10064,
10583..10659,10807..11354)
/ gene="OJ1065_B06.2"
/ note="supported by full-length cDNA(s): AK102382"
/ join(4298..4486,5711..5834,5914..6058,6190..6415,
7021..7094,7268..7360,8691..8808,9598..9660,10032..10064,
10583..10657)
/ gene="OJ1065_B06.2"
/ note="contains full-length cDNA(s):
AK072012, AK102382, AK063129"
/ codon_start=1
/ product="putative tocopherol polyprenyltransferase"
/ protein_id="BAC83059.1"
/ db_xref="GI:34393498"
/ translation="MASLASPLPCRAAATASRSRGPAPRLLGPPPPASPILLSASA
RPPRAANAAWSRDRVAVCVCSQAAGAPALSTLSLDKSCWRFLRPIRTALG
SMILVARALINPOLINWLVKAFYGLVALICNGYIVGINOYDIRIDKNAPYLP
TAAGDLVQTAWLAVLPAAGESIVTNPGPFTSLYCLGLFTGLIYVPEFLKRY
PVAELLIATVRGFLNFGVYATRAALGLTFQWSSPVAFITCFVTLFALVIATKDL
PDVGGRIQISTLATKQGVNRNIAFLGSLLIANYAAIAVAFMLPQAFRTVMVPVH
AALAVGIIFQWVLEQAKYTKDAISQYRFIWNLFYAEYIFFLI"
complement(join(13924..14038,14095..14147,14201..14275))
/ gene="OJ1065_B06.3"
complement(join(13924..14038,14095..14147,14201..14275))
/ gene="OJ1065_B06.3"
/ note="hypothetical ORF
predicted by GeneMark.hmm
this category is not included in IRGSP standard"
16161..20833
/ join(<16161..16612,19790..20147,20207..>20833)
/ gene="OJ1065_B06.4"
/ note="start and end point are not identified"
/ join(16161..16612,19790..20147,20207..20833)
/ gene="OJ1065_B06.4"
/ codon_start=1
/ product="putative Nt-gh3 deduced protein"
/ protein_id="BAC83060.1"
/ db_xref="GI:34393499"
/ translation="MAAGKSTIGTAASRLSSLSMAAKEPDVENRLRIELTGNVDA
VQERVLAIIIGNDAEYIDKGLDASDTDRATFRKVPVASYDDLKPKYKRIANGOR
SPILSTHPIIEFTTSAGERKLMPITVDEMARREVLSLATSVLNVYVPGHGTGK
GLYFLFARSEYKTPGGLTAQALTSYKSEHFKRAYITSMAAILCEDASQSMYAMQ
LCGLCQRHDVLRVGAFFAALVRAIRFLQNLWAOLADTETGLNPRVTDSPDECSR
GDWTGIVTRLPKTKCLNVVTVGWAQYIPITQYVSGGLPIVSGWYASSECFEGLNLR

```

```

PLCGPSEVSYTINPTAYFEFLPVGEAVDASNLVELARVEDGREYEVVVVTVAGLNRY
RGDVLVCTGFHNAQPFRRVQRQSLISLEADKDEAELRAVERASSALLRPRGAS
IVYTSRACTERVPGHFSFALPHWTLP"
complement(join(21777..21922,22075..22156))
/ gene="OJ1065_B06.5"
complement(join(21777..21922,22075..22156))
/ gene="OJ1065_B06.5"
/ note="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
22591..23165
/ gene="OJ1065_B06.6"
/ join(<22591..22692,23007..>23165)
/ gene="OJ1065_B06.6"
/ notes="start and end point are not identified"
/ join(22591..22692,23007..23165)
/ gene="OJ1065_B06.6"
/ notes="predicted by GENSCAN etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83061.1"
/ db_xref="GI:34393500"
/ translation="MSGHGCADCWVMPFCRIGTLCOGRGPAQAARSNSRQTSAMHALL
ASRVAPHTTSPPPPQLTPSPSPQSGERTSSGRSTMAGPG"
complement(23923..24723)
/ gene="OJ1065_B06.7"
complement(join(<23923..24111,24601..>24723))
/ gene="OJ1065_B06.7"
/ notes="start and end point are not identified"
complement(join(23923..24111,24601..24723))
/ gene="OJ1065_B06.7"
/ notes="predicted by FGENSEH etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83062.1"
/ db_xref="GI:34393501"
/ translation="MSAAQLFVLSLFPPLSPSLANDSGSPGMARYRVDGGGTATT
RQVDIDCIHGTTTKRAAGKKAAMQECASPPANYFKNKADYNAAAASASAGAINY
K"
25126..26796
/ gene="OJ1065_B06.8"
/ join(<25126..25183,25708..26045,26134..>26796)
/ gene="OJ1065_B06.8"
/ notes="start and end point are not identified"
/ join(25126..25183,25708..26045,26134..26796)
/ gene="OJ1065_B06.8"
/ note="predicted by GeneMark.hmm etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83063.1"
/ db_xref="GI:34393502"
/ translation="MAQPNDOHNDPTSPSGIROWAQOAGGSGSAAGDGRGD
GGSQHDIPSPDAPKQSTPPTTSPAPSRLOCNHIGAKIKTNSAINGVPVLLLF
VTFVSLGGGGVLLIVEGDDGAAREDEPSLPDPSLRQDVLVLAQIALAVDRA
TDVGIKTACGSPQSAAGAFAMWRESGVDATAVAGATTIDVLSLLSPSPNHLVLP

```

Query Match 89.5%; Score 17; DB 8; Length 113170;
 Best Local Similarity 100.0%; Pred. No. 5.1e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGAC 18
 |||||
 Db 72265 TCGACGCCGCCGATGAC 72249

RESULT 22
 AP003013/c
 LOCUS AP003013 347950 bp DNA linear BCT 15-MAY-2001
 DEFINITION Mesorhizobium loti DNA, complete genome, section 20/21.
 ACCESSION AP003013 BA000012
 VERSION AP003013.2 GI:14027324
 KEYWORDS Mesorhizobium loti
 SOURCE

| | | | |
|-----------|---|--|--|
| ORGANISM | Mesorhizobium loti | | |
| REFERENCE | Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Mesorhizobium. | | |
| AUTHORS | 1 (sites) Kaneko.T., Nakamura.Y., Sato.S., Asamizu.E., Kato.T., Sasamoto.S., Watanabe.A., Ideawara.K., Ishikawa.A., Kawashima.K., Kimura.T., Kishida.Y., Kiyokawa.C., Kohara.M., Matsumoto.M., Matsumoto.A., Mochizuki.Y., Nakayama.S., Nakazaki.N., Shimo.S., Sugimoto.M., Takeuchi.C., Yamada.M. and Tabata.S. | | |
| TITLE | Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti | | |
| JOURNAL | DNA Res. 7 (6), 331-338 (2000) | | |
| MEDLINE | 21082930 | | |
| PUBMED | 11214968 | | |
| REFERENCE | 2 (bases 1 to 347950) | | |
| AUTHORS | Kaneko.T. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (05-DEC-2000) Takakazu Kaneko. Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: kaneko@kazusa.or.jp, /rhizobase/, URL: http://www.kazusa.or.jp/) Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934) On May 11, 2001 this sequence version replaced gi:11994988. | | |
| COMMENT | Location/Qualifiers | | |
| FEATURES | source 1..347950 /organism="Mesorhizobium loti" /mol_type="genomic DNA" /strain="MAFF303099" /db_xref="taxon:381" complement(78..1037) /gene="ml17923" complement(78..1037) /gene="ml17923" /codon_start=1 /transl_table=11 /product="transcriptional regulator" /protein_id="BAB53594.1" /db_xref="GI:14027325" /translation="MRLSQVNLNLSKIVESAARHCNFTPAGERQFIFASVSQVKS LEDQRFKIFQGGNAVLTPETGTVSRREALRIVAASEATGQSEHVLKISVL PTFAARWLPFRPQRIYDPIDVMKVSYSATHEFTSDFDLEIRYGDGHFGLPSDL LFREDLTPVCSKLFHEVLGDKPVSKEPDDLRLHFTLLHSDTCTQNMQSWLGFAGAF VLSFTKSYFDSCMSVEAANAGMGFAVANRAYASDIRAEKLVAPFAVHPHNSAGWY FVSHKCLAAARKVLLFKQWMAEALTQQLDSEIRDLATEAV" 1150..1986 /gene="mlr7924" 1150..1986 /gene="mlr7924" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB53595.1" /db_xref="GI:14027326" /translation="MNAHTIPELRYAMGREAIIGHETAKWVSSFGVAQYLHGYPALL AAIEEALKDKSHAVKHLDLFTFGADRIPEIKELLDKRLERLSDNMGKLEP YPLSTVGTVTMPNPDGAVWHCDGVPTLPLISLGNPLGVGHLEIYCDSDTGBA ILESGREIPRNIRIMIDHKMYATLQGLVGLHRTAPIQYGERVTLVLRQSRVAKPY DNRWFYLAADNDHNDREWVNELAEDVTNQLPAYRFRFAEHPTPAPASETCVSGARG SN" 2032..2958 /gene="mlr7925" 2032..2958 /gene="mlr7925" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB53596.1" /db_xref="GI:14027327" /translation="MVQHRDRAPGRFLFRHGSSSIALVFAAGAVALMSTNALVGSLL ATHVSVQVQLQFAGAAVLIRFMSREAPSAVPAFAAALAAVGLVGTWVLQY IAFASMPTEANLVATWPLMVAIAFIADPNRPALLGLAAVLGFVGLVAVISGGRE NNWFQGLVGYFAAFGSALCMAYFSVMVGRLATSPDRLLPLLSLVGLVALALLWSAGEG | | |
| gene | | VAMPAGMDALGLYLIGAGPMGLGYFWSRALKLEGGKVVAVVAYLTPTIASTLLLTLSG EOLTTTAAIAGAVLIVIGSCIAVGLRSEAQDYV" 2951..3154 /gene="msr7926" 2951..3154 /gene="msr7926" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB53597.1" /db_xref="GI:14027328" /translation="MFDNRLARQEAHWLKEFGAEAPLYAAMKAERAEQKDFGRC ASWRRILEILADARTTKSAKY" 3307..3759 /gene="mlr7927" 3307..3759 /gene="mlr7927" /transl_table=11 /product="transcriptional regulator" /protein_id="BAB53598.1" /db_xref="GI:14027329" /translation="MKELSNIEDKTVIELTADIVAYVGNPLPASGLPDLIASVSAS VRKLAAVGVESPLVAVNPKKSVFPDYIICLEDGKKFKSLKRHLRTDYGSLPDDYR AKWGLPPDYPMVAPNYSATRSALAKSTGLGRKPAAPAAVAKKAKA" 3916..4065 /gene="msr7929" 3916..4065 /gene="msr7929" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB53599.1" /db_xref="GI:14027330" /translation="MNCWAYFEAKGRSALVLDKMFITLFLLELGGSKPLDNQHCRA SURPR" complement(4476..4730) /gene="msl7931" complement(4476..4730) /gene="msl7931" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB53600.1" /db_xref="GI:14027331" /translation="NRSERSLMISMAALLHRTSLTKALRASWPDVKTGGSGSALRH OPGKIKRRSDPDCQVADRKTAGTDIEDTGLNLYFYSC" 4699..6576 /gene="mlr7932" 4699..6576 /gene="mlr7932" /codon_start=1 /transl_table=11 /product="ABC transporter, ATP-binding component" /protein_id="BAB53601.1" /db_xref="GI:14027332" /translation="MLIINDLSRMAGRLLLDHSITLTPAGTKAGLVGRNGTKTTLF KAITGPPETGSIPLPNKTRIGQVAQEAFFGTEELIEVLKADVERTALLBEEKTAT DPHRIADIMRLADIDAHSAEAAATILAGLGFDDAAQGPASSFGSGWRMVAALAAV LFSEPLLLEDEPTNYLDEGLTLENVYVKYPTVLLISHDRDLINRAVNSIVHLDO KLTFWRGYDOPEROYTELQEKQVKAARKHMFVERFRAKAKARQAQSR IKALEKMP2IAAIVDSVZPFSPBPVKTVAIPVALNNVNYTEGQFLLKMTLRI DADDRALLCANGNGKSTAKLLSGRLKQESGTMVAQGLKVAIFAQQLDLRBEEN AYEHRVLRMPAPESKVRGVAFGLTTTMMNTPAKDUSGGEKARLLMGLSAFEPNL FILEDPTNHLIDISRESLHNLNPPGAVILISHDRHLEATADRLWLVKUGAVNPYD GLODEYTLTVGTGSDRGKREADKASADRRAAFAFEPLAKETRAATEALMDR IKRKIDIGIELANPAIYEKDPSTATLAKERSQLAQTLAGHEEKWLTMSAYEBGTA E" 6667..8178 /gene="mlr7933" 6667..8178 /gene="mlr7933" /note="similar to formate dehydrogenase" | |

```

/codon_start=1
/transl_table=11
/protein_id="BAB53602.1"
/db_xref="GI:14027333"
/translation="MNQAKRLRHSHSACPHDCSTCALEVELLDGNRIKRVHGAKANT
YTSVCAKARVADRVHHPDRLLKPLIRAGAKGDGVQWQASWEALDLVAERKFAAE
YKSGEIVPYPFAASCIATVGSQWQEGGAFHNSGIFKLNQEVLEGTRMRDPSIR
YLHRSRTGPVLITGASDALYGGPPVTAMLIONTPNVNAPBOLVKOGFLRDDLDLFACVH
EQMTDARLADVLPAATMEHDDVYKGGNOHITLGPKLIEPPGSPNTHFVIGOL
AERGVADRFGGLTGEQOHLIDILKRGGLGSPSLKBEKQKVDLQBPFEAHFVIGTGH
SDGFRPDMWTGQAANRPKSMGLFEPFPPHDDLLDGLADGRVEVGNQDGLVHAKFF
ARNFLNFTFAETPVSKGEGRPALLHPDDAIDLADGRVEVGNQDGLVHAKFF
DGIRGVVIAEGVWPNSEHERGGINVLTGADAPAPYGGAAFHNDKVMVLEKAG"
complement(8203..9390)
/gene="m117934"
/codon_start=1
/transl_table=11
/protein_id="BAB53603.1"
/db_xref="GI:14027334"
/translation="MFLTSGRELSREARLDDTTTVAIIGAGPAGLAVACLQAGVD
FIIEKQQAAPWRRHYERVHLHTTKRYSLSLPFPVPPKPYPRPRALFVLDIAYIA
QRDLRFQGTAVKAVTODGRGMRVDASGFLRAKDVVIASGYNAPFLRPAFADITF
TGKTLHGADYNNAKLPFAGQSVLITGMNTGAEIALDLAENGAQFTTSVRGGVHIVPRE
LFGVPIQWAGVNAALGPRLNDALFPVILDLVLGRDKYGLKRPQGCGLLEQIAVARSI
PVIDVGTIKIREGAIKVPDITEISORGARFADGKHGFDAIIFATGVPRGVARFLE
PGIQDRSGVTPKASDILGLYLIGHNAVTLGRIEIGIEAQAIAADDIHRHLNRKKAEEI
LPV"
complement(9426..9905)
/gene="m117935"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53604.1"
/db_xref="GI:14027335"
/translation="MPPSPPLDSFDLAILVLQDNTTPQRLIGEAVNLSPAPVQR
IKRQAGVIAASNVAIIEPAAGKPIITIFVEVELESRTLIDAARQFQSTQTEVQCC
YVVTGEADFLIVTADMGAYEALTKLFFGSNNVRKFRITFVAMRVKVLGLTVPLEPD"
10133..11218
/gene="m117936"
10133..11218
/gene="m117936"
/notes="hypothetical protein"
/codon_start=1

Query Match 89.5%; Score 17; DB 1; Length 347950;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17
|||||
Db 198925 GTCGACGCCGCCGATGA 198909

RESULT 23
AY023290/c
LOCUS 227 bp DNA linear PLN 07-FEB-2001
DEFINITION Oryza sativa microsatellite MRG5615 containing (GGC)X9, closest to
marker G148, genomic sequence.
ACCESSION AY023290
VERSION AY023290.1 GI:12706506
KEYWORDS Oryza sativa
SOURCE Oryza sativa
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 227)
REFERENCE
AUTHORS Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.

```

```

TITLE Simple sequence repeats from Monsanto rice genomic sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 227)
AUTHORS Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
Blvd., Creve Coeur, MO 63167, USA
COMMENT Derived from rice genomic sequences generated from the Monsanto
Rice Genome Sequencing project. Please see
http://www.rice-research.org for more information. The sequence
data were produced primarily in the laboratories of Dr. Leroy Hood
at the University of Washington in Seattle.
FEATURES
source
1..227
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4330"
repeat_region
1..227
/note="microsatellite MRG5615"
/rpt_type=tandem
/rpt_unit="ggc"
ORIGIN
Query Match 86.3%; Score 16.4; DB 8; Length 227;
Best Local Similarity 94.4%; Pred. No. 1.4e-04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGAC 18
|||||
Db 41 GTCGACGCCGCCGAGGAC 24

RESULT 24
AFAAZU/c
LOCUS 810 bp DNA linear BCT 24-APR-1993
DEFINITION A. denitrificans azurin (azu) Gene, complete cds.
ACCESSION X30388
VERSION X30388.1 GI:141901
KEYWORDS azurin.
SOURCE Achromobacter xylosoxidans subsp. denitrificans
ORGANISM Achromobacter xylosoxidans subsp. denitrificans
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Achromobacter.
1 (bases 1 to 810)
REFERENCE
AUTHORS Hoitink,C.W., Woudt,L.P., Turenhout,J.C., van de Kamp,M. and
Canter,G.W.
TITLE Isolation and sequencing of the Alcaligenes denitrificans
azurin-encoding gene: comparison with the genes encoding blue
copper proteins from Pseudomonas aeruginosa and Alcaligenes
faecalis
JOURNAL Gene 90 (1), 15-20 (1990)
MEDLINE 9037337
PUBMED 2116366
COMMENT Original source text: A.denitrificans (strain NCTC9582) DNA.
Draft entry and computer-readable sequence for (1) kindly submitted
by G.W.Canter, 01-DEC-1989, for release after publication.
FEATURES
source
1..810
/organism="Achromobacter xylosoxidans subsp.
denitrificans"
/mol_type="genomic DNA"
/sub_species="denitrificans"
/db_xref="taxon:32002"
190..203
/notes="fmr-box"
251..267
/notes="ntrA-box"
296..299
/notes="ribosome binding site"
307..756
/notes="azurin (azu) precursor"
/codon_start=1
/transl_table=11

```

```

/protein_id="AA21954.1"
/db_xref="GI:141902"
/translation="MLAKATLAIVLSAASLPVLAQCEATIESNDAMQYNLKENWVDK
SKQFTVHLKHVKVAKVANGHNVLTREADQGVATDGMNAGLAQDYYKAGDTRVIA
HTKVIIGGESDSVTDFVSKLTPGEAYAFYFCFPGHWAMKGTGLKLSN"
sig_peptide
307..366
/note="azurin signal peptide"
mat_peptide
367..753
/product="azurin"
misc_signal
778..800
/note="terminator"

ORIGIN
Query Match 86.3%; Score 16.4; DB 1; Length 810;
Best Local Similarity 94.4%; Pred. No. 1.3e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19
Db 638 TCGCGCGCGCGATGACC 621

RESULT 25
AX653785
LOCUS AX653785/c
DEFINITION Sequence 3655 from Patent WO03000898.
ACCESSION AX653785
VERSION AX653785.1 GI:29156599
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Katagiri,F., Quan,S., Tao,X., Whitcham,S., Xie,Z., Zhu,T. and Zou,G.
Plant genes involved in defense against pathogens
Patent: WO 03000898-A 3655 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
Location/Qualifiers
source
1..963
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 86.3%; Score 16.4; DB 6; Length 963;
Best Local Similarity 94.4%; Pred. No. 1.3e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGAC 18
Db 501 GTCGCGCGCGCGATGAC 484

RESULT 26
AX660080/c
LOCUS AX660080/c
DEFINITION Sequence 437 from Patent WO03000906.
ACCESSION AX660080
VERSION AX660080.1 GI:29162039
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 Glazebrook,J., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
Katagiri,F., Kreps,J., Provart,N., Rieke,D. and Zhu,T.
Plant disease resistance genes
Patent: WO 03000906-A 437 03-JAN-2003;

```

```

Syngenta Participations AG (CH)
FEATURES
source
1..1077
Location/Qualifiers
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 86.3%; Score 16.4; DB 6; Length 1077;
Best Local Similarity 94.4%; Pred. No. 1.3e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGAC 18
Db 171 GTCGACGTCGCGATGAC 154

RESULT 27
AB025282/c
LOCUS AB025282
DEFINITION Gallus gallus cpgc mRNA for pepsinogen C, complete cds.
ACCESSION AB025282
VERSION AB025282.1 GI:4589839
KEYWORDS pepsinogen C.
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (sites)
AUTHORS Sakamoto,N., Saiga,H. and Yasugi,S.
TITLE Analysis of temporal expression pattern and cis-regulatory
sequences of chicken pepsinogen A and C
JOURNAL Biochem. Biophys. Res. Commun. 250 (2), 420-424 (1998)
MEDLINE 98440813
PubMed 9753645
REFERENCE
2 (bases 1 to 1339)
AUTHORS Sakamoto,N. and Yasugi,S.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Sadao Yasugi, Tokyo Metropolitan
University, Department of Biology, The Graduate School of Science;
Minamiohsawa 1-1, Hachioji, Tokyo 192-0397, Japan
(E-mail: yasugi-sadao@ec.metro-u.ac.jp, Tel: +81-426-77-2572,
Fax: +81-426-77-2559)
FEATURES
Location/Qualifiers
source
1..1339
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
1..1339
/gene="cPgC"
20..1189
/gene="cPgC"
/codon_start=1
/product="pepsinogen C"
/protein_id="BAA76893.1"
/db_xref="GI:4589840"
/translation="MKRLILTVLCLHLCGILRVPLKKGKSTREAKESGVLDHYLAN
HHRYDPAYKPFNFATAYEPLANNDMSYGEISIGTPQNFVLVLDTGSSNLWVEST
LQSQCAKNHNFDFNESTFTQDFFSLQSGSLTGIFGFDVDTVIGQISITNQEF
GLSETEPGTFLYSPFDGLLAFPSISAGGATVQMKQLNLLDFPVFSYLSGQE
GSGGGLVFGVDPLNYLTQITVPTQTYWQIGIEDFVAGGQSGWCSQCQCIDV
TGTSLTVFNVQVFTLMQYIGAADSDGQYVASCNIEVMTITFVIGTSFPLPPSA
YMLQNSNDYCTVGIESTYLPSTGQPLWILGDVFLRVYSYIDMGNNGVFATAV"
1339
/gene="cPgC"
/note="9 A nucleotides"

polyA_site
1339

ORIGIN
Query Match 86.3%; Score 16.4; DB 5; Length 1339;
Best Local Similarity 94.4%; Pred. No. 1.2e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```



```

Qy 1 GTCGACGCCGCCGATGAC 18
|||||
Db 712 GTCGACGCCGCCGAGAC 595

RESULT 28
AK104317/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) cDNA clone:001-024-C03, full
insert sequence.
ACCESSION
AK104317
VERSION
AK104317.1 GI:32989526
FLI CDNA; oligo capping.
KEYWORDS
Oriza sativa (japonica cultivar-group)
SOURCE
Oriza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1
The Rice Full-Length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-Length cDNA Project Team,
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I.,
Hirashida, K., Hayashi, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Kojima, K., Namiki, T., Ohtsuka, T., Carninci, P., Carninci, P.,
Ohtsuka, T., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ohtsuka, T., Iida, Y.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
japonica rice
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)
2752273
12869764
2 (bases 1 to 1389)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Hayashi, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirashida, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kawagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuka, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ohtsuka, T., Ohtsuka, T., Ohtsuka, T., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, P., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and
Yoshimura, A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agricultural Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan [E-mail: skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007]
This clone is one of the 28k full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,

```

```

Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., and
Yamamoto, M.
FAIS Genome Sequencing & Analysis Group: Ohtsuka, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K., and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hirashida, K., Iida, J., Iida, Y., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Oka, M., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, P.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A., and Hayashizaki, Y.
Location/Qualifiers
1..1389
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-024-C03"

FEATURES
source
1..1389
Query Match 86.3%; Score 16.4; DB 8; Length 1389;
Best Local Similarity 94.4%; Pred. No. 1.2e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
|||||
Db 608 GTCGACGCCGCCGATGAC 591

RESULT 29
AK071251/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) cDNA clone:J023087H21, full
insert sequence.
ACCESSION
AK071251
VERSION
AK071251.1 GI:32981274
FLI CDNA; CAP trapper.
KEYWORDS
Oriza sativa (japonica cultivar-group)
SOURCE
Oriza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1
The Rice Full-length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-length cDNA Project Team,
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I.,
Hirashida, K., Hayashi, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Kojima, K., Namiki, T., Ohtsuka, T., Carninci, P., Carninci, P.,
Ohtsuka, T., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Ohtsuka, Y., Iida, Y.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12869764
2 (bases 1 to 1502)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imoto,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kadowa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kuroaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Nariakawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,B., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otsu,Y., Otsu,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Sato,K., Shibata,K., Shikata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yanagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kanondai, Tsukuba, Ibaraki 305-8602, Japan E-mail:skkikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007
This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Sato,H., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yanagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Nariakawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imoto,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Otsu,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
Location/Qualifiers
1..1502
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J023087H21"

TITLE
JOURNAL

Query Match 86.3%; Score 16.4; DB 8; Length 1502;
Best Local Similarity 94.4%; Pred. No. 1.2e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

FEATURES
source
Location/Qualifiers
1..1560
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 86.3%; Score 16.4; DB 6; Length 1560;
Best Local Similarity 94.4%; Pred. No. 1.2e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGACGCCGCCGATGAC 18
|||||
Db 624 GTGACGCCGCCGATGAC 607

RESULT 30
AX366999/c
LOCUS AX366999 1560 bp DNA linear PAT 16-FEB-2002
DEFINITION Sequence 26 from Patent WO0198509.
ACCESSION AX366999
VERSION AX366999.1 GI:18698275
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
1
REFERENCE
AUTHORS Lanahan,M.B., Desai,N.M. and Gasdaska,P.Y.
TITLE Grain processing method and transgenic plants useful therein
JOURNAL Patent: WO 0198509-A 26 27-DEC-2001;
SYNOPSIS Syngenta Participations AG (CH)
FEATURES
source
Location/Qualifiers
1..1560
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

Qy 1 GTGACGCCGCCGATGAC 18
|||||
Db 906 GTGACGCCGCCGATGAC 889

Search completed: June 20, 2004, 11:43:55
Job time : 531.915 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 678.836 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccggagggccagagggacatagc 25

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.btg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rtd.*

36: em.htg.man.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 25 | 100.0 | 15619 | 1 | AE006916 | AE006916 Mycobacte |
| 2 | 25 | 100.0 | 341957 | 15 | BX842572 | BX842572 Mycobacte |
| 3 | 25 | 100.0 | 343050 | 1 | BX248334 | BX248334 Mycobacte |
| 4 | 24 | 96.0 | 38721 | 1 | MSGY219 | AD000013 Mycobacte |
| 5 | 20.2 | 80.8 | 60740 | 2 | AC016557 | AC016557 Homo sapi |
| 6 | 20.2 | 80.8 | 168008 | 9 | AC008544 | AC008544 Homo sapi |
| 7 | 20.2 | 80.8 | 171004 | 2 | AC034202 | AC034202 Homo sapi |
| 8 | 20.2 | 80.8 | 183734 | 9 | AC109479 | AC109479 Homo sapi |
| 9 | 20.2 | 80.8 | 185329 | 2 | AC023587 | AC023587 Homo sapi |
| 10 | 20.2 | 80.8 | 199768 | 2 | AC136594 | AC136594 Homo sapi |
| 11 | 20.2 | 80.8 | 220632 | 2 | AC136596 | AC136596 Homo sapi |
| 12 | 19.2 | 76.8 | 104526 | 9 | AC114743 | AC114743 Homo sapi |
| 13 | 19.2 | 76.8 | 146868 | 2 | AC016802 | AC016802 Homo sapi |
| 14 | 19.2 | 76.8 | 166558 | 2 | AC112141 | AC112141 Homo sapi |
| 15 | 19.2 | 76.8 | 169651 | 2 | AC084771 | AC084771 Homo sapi |
| 16 | 18.6 | 75.2 | 258723 | 2 | AC110113 | AC110113 Rattus no |
| 17 | 18.6 | 74.4 | 1276 | 10 | WU93191 | U93191 Mus musculu |
| 18 | 18.6 | 74.4 | 59817 | 6 | AX695375 | AX695375 Sequence |
| 19 | 18.6 | 74.4 | 60042 | 2 | AC100484 | AC100484 Mus muscu |
| 20 | 18.6 | 74.4 | 92588 | 9 | HSIMDL | X97051 DNA sequenc |
| 21 | 18.6 | 74.4 | 100057 | 10 | AL929178 | AL929178 Mouse DNA |
| 22 | 18.6 | 74.4 | 145052 | 10 | AL603389 | AL603389 Mouse DNA |
| 23 | 18.6 | 74.4 | 149321 | 2 | AL731845 | AL731845 Mus muscu |
| 24 | 18.6 | 74.4 | 151555 | 10 | AL732474 | AL732474 Mouse DNA |
| 25 | 18.6 | 74.4 | 157090 | 9 | AB019441 | AB019441 Homo sapi |
| 26 | 18.6 | 74.4 | 157361 | 10 | AL645923 | AL645923 Mouse DNA |
| 27 | 18.6 | 74.4 | 165116 | 4 | AC092729 | AC092729 Canis fam |
| 28 | 18.6 | 74.4 | 165588 | 9 | AL590966 | AL590966 Human DNA |
| 29 | 18.6 | 74.4 | 169802 | 9 | CNS01DT2 | AL121217 Human chr |
| 30 | 18.6 | 74.4 | 183564 | 10 | BX546505 | BX546505 Mouse DNA |
| 31 | 18.6 | 74.4 | 183786 | 10 | AL607087 | AL607087 Mouse DNA |
| 32 | 18.6 | 74.4 | 190512 | 2 | AC140278 | AC140278 Mus muscu |
| 33 | 18.6 | 74.4 | 190981 | 2 | BX890380 | BX890380 Mus muscu |
| 34 | 18.6 | 74.4 | 191941 | 2 | AC102693 | AC102693 Mus muscu |
| 35 | 18.6 | 74.4 | 191969 | 2 | AC101974 | AC101974 Mus muscu |
| 36 | 18.6 | 74.4 | 195848 | 2 | AC102615 | AC102615 Mus muscu |
| 37 | 18.6 | 74.4 | 200088 | 10 | AC078897 | AC078897 Mus muscu |
| 38 | 18.6 | 74.4 | 202848 | 2 | AC113503 | AC113503 Mus muscu |
| 39 | 18.6 | 74.4 | 203592 | 10 | AL669921 | AL669921 Mouse DNA |
| 40 | 18.6 | 74.4 | 210369 | 2 | AC109502 | AC109502 Mus muscu |
| 41 | 18.6 | 74.4 | 210919 | 2 | AC115554 | AC115554 Rattus no |
| 42 | 18.6 | 74.4 | 232278 | 2 | AC087876 | AC087876 Mus muscu |
| 43 | 18.6 | 74.4 | 237282 | 10 | AL772371 | AL772371 Mouse DNA |
| 44 | 18.6 | 74.4 | 239525 | 2 | AC095150 | AC095150 Rattus no |
| 45 | 18.6 | 74.4 | 241082 | 2 | AC127158 | AC127158 Rattus no |
| 46 | 18.6 | 74.4 | 242224 | 2 | AC105066 | AC105066 Mus muscu |
| 47 | 18.6 | 74.4 | 243659 | 2 | AC128597 | AC128597 Rattus no |
| 48 | 18.6 | 74.4 | 244903 | 2 | AC097807 | AC097807 Rattus no |
| 49 | 18.6 | 74.4 | 257471 | 2 | AC139675 | AC139675 Mus muscu |
| 50 | 18.6 | 74.4 | 267943 | 2 | AC136702 | AC136702 Mus muscu |
| 51 | 18.6 | 74.4 | 309350 | 1 | AP005215 | AP005215 Corynebact |
| 52 | 18.6 | 74.4 | 309304 | 1 | AE016865 | AE016865 Pseudomon |
| 53 | 18.6 | 73.6 | 10955 | 9 | HMMHRC01 | M23903 Human MHC C |
| 54 | 18.4 | 73.6 | 14646 | 9 | HSHLASBA | X03100 Human HLA-S |
| 55 | 18.4 | 73.6 | 14782 | 9 | HSHLADFB | X02228 Human HLA-D |
| 56 | 18.4 | 73.6 | 64380 | 9 | BX120009 | BX120009 Human DNA |
| 57 | 18.4 | 73.6 | 106728 | 9 | AL805913 | AL805913 Human DNA |
| 58 | 18.4 | 73.6 | 124899 | 9 | AL645931 | AL645931 Human DNA |
| 59 | 18.4 | 73.6 | 181228 | 2 | AC011086 | AC011086 Homo sapi |
| 60 | 18.4 | 73.6 | 187964 | 9 | AL662824 | AL662824 Human DNA |
| 61 | 18.4 | 73.6 | 196572 | 2 | AC118785 | AC118785 Rattus no |
| 62 | 18.4 | 73.6 | 243279 | 2 | AC099068 | AC099068 Rattus no |
| 63 | 18.4 | 73.6 | 272405 | 2 | AC094363 | AC094363 Rattus no |
| 64 | 18.2 | 72.8 | 1397 | 3 | AY069277 | AY069277 Drosophil |
| 65 | 18.2 | 72.8 | 2140 | 3 | DMCOLT | Y12495 D.melanogas |

Pred. No. is the number of results predicted by chance to have a

| | | | | | | |
|-------|------|------|--------|----|------------|--------------------|
| C 212 | 17.6 | 70.4 | 205619 | 2 | AC108275 | Rattus no |
| C 213 | 17.6 | 70.4 | 208406 | 2 | AC133405 | Rattus no |
| C 214 | 17.6 | 70.4 | 209378 | 2 | AC137262 | Rattus no |
| C 215 | 17.6 | 70.4 | 209451 | 2 | AC124942 | Rattus no |
| C 216 | 17.6 | 70.4 | 216825 | 2 | AC108908 | Mus muscu |
| C 217 | 17.6 | 70.4 | 217595 | 10 | AL805959 | Mouse DNA |
| C 218 | 17.6 | 70.4 | 219315 | 2 | AC119914 | Mus muscu |
| C 219 | 17.6 | 70.4 | 219548 | 2 | AC034265 | Mus muscu |
| C 220 | 17.6 | 70.4 | 221154 | 2 | AC111543 | Rattus no |
| C 221 | 17.6 | 70.4 | 221968 | 2 | AC115188 | Rattus no |
| C 222 | 17.6 | 70.4 | 224973 | 2 | AC118082 | Rattus no |
| C 223 | 17.6 | 70.4 | 233042 | 10 | AC132144 | Mus muscu |
| C 224 | 17.6 | 70.4 | 234429 | 9 | AF311103 | Homo sapi |
| C 225 | 17.6 | 70.4 | 235978 | 2 | AC094420 | Rattus no |
| C 226 | 17.6 | 70.4 | 236760 | 2 | AC132537 | Rattus no |
| C 227 | 17.6 | 70.4 | 237644 | 2 | AC095758 | Rattus no |
| C 228 | 17.6 | 70.4 | 237913 | 2 | AC094576 | Rattus no |
| C 229 | 17.6 | 70.4 | 239486 | 2 | AC094381 | Rattus no |
| C 230 | 17.6 | 70.4 | 240657 | 2 | AC111675 | Rattus no |
| C 231 | 17.6 | 70.4 | 242507 | 2 | AC134246 | Mus muscu |
| C 232 | 17.6 | 70.4 | 242825 | 9 | HSPEX | Y10196 Homo sapien |
| C 233 | 17.6 | 70.4 | 244376 | 2 | AC120968 | Rattus no |
| C 234 | 17.6 | 70.4 | 246688 | 2 | AC112282 | Rattus no |
| C 235 | 17.6 | 70.4 | 247504 | 2 | AC101851 | Mus muscu |
| C 236 | 17.6 | 70.4 | 249206 | 2 | AC145168 | Mus muscu |
| C 237 | 17.6 | 70.4 | 252175 | 2 | AC107539 | Rattus no |
| C 238 | 17.6 | 70.4 | 252283 | 2 | AC126977 | Rattus no |
| C 239 | 17.6 | 70.4 | 252348 | 2 | AC094064 | Rattus no |
| C 240 | 17.6 | 70.4 | 253409 | 2 | AC106926 | Rattus no |
| C 241 | 17.6 | 70.4 | 253819 | 2 | AC131299 | Mus muscu |
| C 242 | 17.6 | 70.4 | 255870 | 2 | AC106968 | Rattus no |
| C 243 | 17.6 | 70.4 | 260164 | 2 | AC098181 | Rattus no |
| C 244 | 17.6 | 70.4 | 260589 | 2 | AC126862 | Rattus no |
| C 245 | 17.6 | 70.4 | 263069 | 2 | AC103440 | Rattus no |
| C 246 | 17.6 | 70.4 | 263388 | 2 | AC094260 | Rattus no |
| C 247 | 17.6 | 70.4 | 273663 | 2 | AC133292 | Rattus no |
| C 248 | 17.6 | 70.4 | 275060 | 2 | AC135406 | Rattus no |
| C 249 | 17.6 | 70.4 | 275924 | 2 | AC116233 | Rattus no |
| C 250 | 17.6 | 70.4 | 303438 | 1 | AE017212 | Geobacter |
| C 251 | 17.6 | 70.4 | 305225 | 2 | AC130763 | Rattus no |
| C 252 | 17.6 | 70.4 | 312905 | 2 | AC094521 | Rattus no |
| C 253 | 17.6 | 70.4 | 347800 | 1 | AP000060 | Aeropyrum |
| C 254 | 17.4 | 69.6 | 671 | 11 | BV014309 | BV014309 S212P6736 |
| C 255 | 17.4 | 69.6 | 36235 | 3 | LMPL7836 | Continuation (4 of |
| C 256 | 17.4 | 69.6 | 110000 | 2 | AC109500_3 | Continuation (3 of |
| C 257 | 17.4 | 69.6 | 110000 | 2 | AC120598_2 | Continuation (4 of |
| C 258 | 17.4 | 69.6 | 117000 | 9 | AC069222 | AC069222 Homo sapi |
| C 259 | 17.4 | 69.6 | 124102 | 9 | AC021148 | AC021148 Homo sapi |
| C 260 | 17.4 | 69.6 | 126589 | 9 | AL512444 | AL512444 Human DNA |
| C 261 | 17.4 | 69.6 | 133289 | 9 | AC012317 | AC012317 Homo sapi |
| C 262 | 17.4 | 69.6 | 148179 | 9 | AP001831 | AP001831 Homo sapi |
| C 263 | 17.4 | 69.6 | 150522 | 9 | AC008870 | AC008870 Homo sapi |
| C 264 | 17.4 | 69.6 | 153024 | 2 | AC133073 | AC133073 Lemur cat |
| C 265 | 17.4 | 69.6 | 156604 | 2 | AL390194 | AL390194 Homo sapi |
| C 266 | 17.4 | 69.6 | 158302 | 2 | AC069481 | AC069481 Homo sapi |
| C 267 | 17.4 | 69.6 | 160212 | 2 | AC009043 | AC009043 Homo sapi |
| C 268 | 17.4 | 69.6 | 163962 | 9 | BS000233 | BS000233 Pan trogl |
| C 269 | 17.4 | 69.6 | 164736 | 9 | AC105275 | AC105275 Homo sapi |
| C 270 | 17.4 | 69.6 | 166992 | 2 | AC022883 | AC022883 Homo sapi |
| C 271 | 17.4 | 69.6 | 171553 | 2 | AC013611 | AC013611 Homo sapi |
| C 272 | 17.4 | 69.6 | 178967 | 2 | AC128772 | AC128772 Rattus no |
| C 273 | 17.4 | 69.6 | 183692 | 8 | AP003230 | AP003230 Oryza sat |
| C 274 | 17.4 | 69.6 | 183988 | 2 | AC012185 | AC012185 Homo sapi |
| C 275 | 17.4 | 69.6 | 188292 | 2 | AC103360 | AC103360 Mus muscu |
| C 276 | 17.4 | 69.6 | 196086 | 10 | AC104889 | AC104889 Mus muscu |
| C 277 | 17.4 | 69.6 | 197909 | 10 | AL928719 | AL928719 Mouse DNA |
| C 278 | 17.4 | 69.6 | 204239 | 10 | AC114819 | AC114819 Mus muscu |
| C 279 | 17.4 | 69.6 | 204693 | 2 | AC131733 | AC131733 Mus muscu |
| C 280 | 17.4 | 69.6 | 211456 | 10 | AL928926 | AL928926 Mouse DNA |
| C 281 | 17.4 | 69.6 | 217521 | 2 | AC016047 | AC016047 Homo sapi |
| C 282 | 17.4 | 69.6 | 219251 | 2 | AC110939 | AC110939 Rattus no |
| C 283 | 17.4 | 69.6 | 219892 | 2 | AC127200 | AC127200 Rattus no |
| C 284 | 17.4 | 69.6 | 227141 | 2 | AC109109 | AC109109 Rattus no |

ALIGNMENTS

RESULT 1
LOCUS AE006916
DEFINITION Mycobacterium tuberculosis CDC1551, section 2 of 280 of the complete genome.
ACCESSION AE006916
VERSION AE006916.1
KEYWORDS Mycobacterium tuberculosis CDC1551
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 15619)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 15619)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 15619
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/notes="clinical strain"
78. .866
/gene="MTC0015"
78. .866
/gene="MTC0015"
/gene="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4237.1"
/db_xref="GI:13879056"
/translation="MELTHTPCPENGEMTIDRRSARFSPVLCVLLAGLIAATHG
VSGGTIRSDAPRLVDLYRQAASVNRILATERALTTRIDSVHGVSVDTAAMQER
SAGLGAANNPVHGGVGLVVLQDAORDANGRPDRSDDLLVHHQDDLEAVLNWN
AGAETQMOPQRIIAISARCVCNTLLNGRTYSPPTTAAIGADAAAMQAAAAAPLV

C 285 17.4 69.6 228894 2 AC133735
C 286 17.4 69.6 229350 2 AC107166
C 287 17.4 69.6 233220 2 AC098058
C 288 17.4 69.6 236716 2 AC098920
C 289 17.4 69.6 267088 2 AC129736
C 290 17.4 69.6 269863 2 AC131017
C 291 17.4 69.6 275605 2 AC123100
C 292 17.4 69.6 278166 2 AC094117
C 293 17.4 69.6 288483 2 AL627109
C 294 17.4 69.6 340000 9 AF001748
C 295 17.2 68.8 461 9 AY027915S2
C 296 17.2 68.8 529 11 BV070169
C 297 17.2 68.8 640 11 G96635
C 298 17.2 68.8 1126 4 PIGGPROAI
C 299 17.2 68.8 2000 6 AX656630
C 300 17.2 68.8 3014 4 AY267904

FT BACTERIA. THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY. THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-INDEPENDENT) ALONG DUPLEX DNA [CATALYTIC ACTIVITY: N deoxynucleoside triphosphate + N diphosphate + {DNA}N].

FT /gene="dna"

FT /locus_tag="Rv0002"

FT /product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA NUCLEOTIDYLTRANSFERASE)"

FT /protein_id="CAA16239.1"

FT /translation="MDAATTRVGLTDLFRLLRESFADAVSWAKNLPARPAVFLVSGV LITGDSNGLTISGFVYEAQVAGAEIVSPGSLVSGRLISDITRALPNKRPVDVHVEG NRVALTCGNARSLPTMPVEDYPTLPTPESTGLLPAELFAEAISQVAILAAGRDITLPM LTGIVEILGEPVLAADRERLAVRELKMSASPDIEAANLVPAKTLAEAKAGIGS DVLISLGTGFGVCKGLGIGSGNKRSTRLIDREFPKFRQLLTETHTAVATMVABELI EAKILVALDAGAGVQRMEFFADGSRVLSAGADDVGRAEEDLVVDYAGEPLTIANPILY TLGSLSLRSERVSFGFTTAGPALLRPVSGDPRPVAGLNGNGPPFPAVSTDYVLLMPVR LFG"

FT CDS

FT 3280..4437

FT /evidence=EXPERIMENTAL

FT /note="Rv0003, (MTCY10H4.01), len: 385 aa, recF, DNA replication and repair protein (see citations below), equivalent to others Mycobacterium DNA replication and repair proteins e.g. NP_301131.1|NC_002677 from Mycobacterium leprae (385 aa); Q9L7L5|REFC_MYCPA from Mycobacterium avium subsp. paratuberculosis (385 aa); P50916|REFC_MYCSM from Mycobacterium smegmatis (384 aa); etc. Also highly similar to others e.g. P36176|REFC_STRCO DNA REPLICATION AND REPAIR PROTEIN from Streptomyces coelicolor (373 aa); NP_440892.1|NC_000911 from Synechocystis sp. strain PCC 6803 (384 aa); NP_469352.1|NC_003212 from Listeria innocua (370 aa); etc. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00617 RecF protein signature 1, and PS00618 RecF protein signature 2. BELONGS TO THE REC F FAMILY."

FT /transl_table=11

FT /function="THE REC F PROTEIN IS INVOLVED IN DNA METABOLISM AND RECOMBINATION; IT IS REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDUCIBILITY. REC F BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS TO BIND ATP."

FT /gene="recF"

FT /locus_tag="Rv0003"

FT /product="DNA REPLICATION AND REPAIR PROTEIN REC F (SINGLE-STRAND DNA BINDING PROTEIN)"

FT /protein_id="CAB02424.1"

FT /translation="MYVRLGLRDPDSRWACVDLELHPGRTVFGNGYKTNLIEALWY SFTLGSRVADPLIRVGTDRVISTVNDGRECAVDLEIATGRVNAKRLNRSSVRS TRDVVGLVRLAFEDLGLVRGDPADRRYLDLAI VRRPAIAVRAEYERVLVORTA LLKSVPGARYGRCGVFTLWVDSRLAEGELVAARI DLAVNQLAPEVKAYCOLLAPE SRASIGTRASNDVGTGPEQSDIDQLLAALLAALARRDAELERGVLPGRDDLI LRLGDPARGHGEASLVAUQLAAYQLLRVDGPEPVLDDVFAELDDVMRRALA TAESAQVLTVAALVLEIDIPAGWDARRVHIDVADDTGMSVLP"

FT misc_feature

FT 3367..3390

FT /note="PS00017 ATP/GTP-binding site motif A"

FT 3634..3690

FT /note="PS00617 RecF protein signature 1"

FT 4243..4296

FT /note="PS00618 RecF protein signature 2"

FT 4434..4497

FT /evidence=EXPERIMENTAL

FT /note="Rv0004, (MTCY10H4.02), len: 187 aa. Conserved hypothetical protein (see Salazar et al., 1996), highly similar, but longer 21 aa in N-terminus, to AAF33696.1|AF222789 unknown protein from Mycobacterium avium subsp. paratuberculosis (166 aa); and highly similar to NP_301132.1|NC_002677 conserved hypothetical protein from Mycobacterium leprae (189 aa); S70990 hypothetical protein from Mycobacterium smegmatis (194 aa). Also highly similar, except in N-terminal part, to

FT Query Match 100.0%; Score 25; DB 15; Length 341957;

FT Best Local Similarity 100.0%; Pred. No. 0.76;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGCGAGGCGCCAGGCGCACATAGC 25
 |||||
 Db 24426 CCGCGAGGCGCCAGGCGCACATAGC 24450
 |||||

RESULT 3
 LOCUS BX248334 343050 bp DNA linear BCT 11-JUN-2003
 DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 1/14.
 ACCESSION BX248334 BX248333
 VERSION BX248334.1 GI:31616762
 KEYWORDS complete genome.
 SOURCE Mycobacterium bovis subsp. bovis AF2122/97
 ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1
 AUTHORS Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, J., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and Hensin, G.
 TITLE The complete genome sequence of Mycobacterium bovis
 JOURNAL Online Publication
 REMARK PNAS 10.1073/pnas.1130426100 (Microbiology)
 REFERENCE 2 (bases 1 to 343050)
 AUTHORS Garnier, T.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex 15, France. e-mail: tgarnier@pasteur.fr Submitted on behalf of the Mycobacterium bovis sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. P4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
 JOURNAL Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES
 source
 1..343050
 /organism="Mycobacterium bovis subsp. bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 1..1524
 /gene="dnaA"
 /locus_tag="Mb0001"
 1..1524
 /locus_tag="dnaA"
 /locus_tag="Mb0001"
 /note="Mb0001, dnaA, len: 507 aa. Equivalent to Rv0001, len: 507 aa, from Mycobacterium tuberculosis strain H37Rv, (99.6% identity in 507 aa overlap). dnaA, chromosomal replication initiator protein (see citations below), equivalent to other Mycobacterial CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g. P46388|DNA_MYCLE from Mycobacterium leprae (502 aa); Q9L7L7|DNA_MYCPA from Mycobacterium paratuberculosis (509 aa); P49990|DNA_MYCAV from Mycobacterium avium (508 aa); P49992|DNA_MYCSM from Mycobacterium smegmatis (504 aa); etc. Also highly similar to others except in N-terminus e.g. Q9ZH75|DNA_STRCH CHROMOSOMAL REPLICATION INITIATOR PROTEIN from Streptomyces chrysomallus (624 aa); Q9ZH76|DNA_STRSE from Streptomyces reticuli (643 aa); DNA_ECOLI|P03004|B3702 chromosomal replication initiator protein from Escherichia coli strain K12 (467 aa), FASTA scores: opt: 986, E(): 0, (43.2% identity in 389 aa overlap); etc. Contains PS00017 ATP/GTP-binding site motif A (P-loop), etc. PS01008 DnaA protein signature. BELONGS TO THE DNAA FAMILY. Note that


```

6803 (384 aa); NP_469352.1|NC_003212 from Listeria innocua
(370 aa); etc. Contains PS00017 ATP/GTP-binding site motif
A (P-loop), PS00617 RecF protein signature 1, and PS00618
RecF protein signature 2. BELONGS TO THE REC_F FAMILY."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="DNA REPLICATION AND REPAIR PROTEIN REC_F
(SINGLE-STRAND DNA BINDING PROTEIN)"
/protein_id="CAD92865.1"
/db_xref="GI:31616765"
/translation="MYVRLHGLRDFRSWACVLELHPGRTVFPVNGYGTKNLILAEI
YSTTGSHRVSAFLRVLGVGTDRIVSTIVNDGRECAVDLEIATGRVKNALNRSSV
RSTDDVGVLLFAPDLRGVDPADRRVLDLAIVRPAPAAVAERVAERVVRVQ
RTALLKSPGAEVGRDGVDFDTLEVWDSLEAEGAEELVAARDLNNQLAEKVAQVL
LAPERSASIGYRSMNDVTGSPSESDTRQLLAARLLANLAARRDELBERGCVLGVPH
RDLILIRGDPKAFSGHSEWSLAVALAQILRRVDGGEPVLLDDVPAELDDUM
RRRLATAESEQVLVTAAVLIEDIPAGWDARRVIDRADDTGSMVSLP"
4434..4997
/locus_tag="Mb00004"
4434..4997
/locus_tag="Mb00004"
/locus_tag="Mb00004"
/notes="Mb00004, -, len: 187 aa. Equivalent to Rv0004, len:
187 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.5% identity in 187 aa overlap). Conserved hypothetical
protein (see citation below), highly similar, but longer
21 aa in N-terminus, to AAF33696.1|AF222789 unknown
protein from Mycobacterium avium subsp. paratuberculosis
(166 aa); and highly similar to NP_301132.1|NC_002677
conserved hypothetical protein from Mycobacterium leprae
(189 aa); S70990 hypothetical protein from Mycobacterium
smegmatis (194 aa). Also similar to C-terminus to
C-terminal part of P35925|YREG_STRCO HYPOTHETICAL 19.8 KDA
PROTEIN (IN REC_F-GYR_B INTERGENIC REGION) from Streptomyces
coelicolor (190 aa), FASTA scores: opt: 404, E(): 3.9e-18,
(40.7% identity in 189 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD92866.1"
/db_xref="GI:31616766"
/translation="MTGSDVDRPDQNGERLMKSPGDLVRLTDEAARAARAGQDAG
RERVASVAGSRVAGRRSSNGSPDIRDPOPLGKAARELAKKESGVRVAEGMVLGVQ
SAVGHQIAEHARPTALNDGVLSVIAESATWATQLRMQAQLLAKAAAVGNDVVRSL
KITGPAPSPWRKGRPHIAGRGPRDTYG"
5123..7267
/genes="gyrB"
/locus_tag="Mb00005"
5123..7267
/genes="gyrB"
/locus_tag="Mb00005"
/SC_number="5.99.1.3"
/notes="Mb00005, gyrB, len: 714 aa. Equivalent to Rv0005,
len: 714 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 714 aa overlap). gyrB, DNA gyrase
subunit B (EC 5.99.1.3) (see citations below), equivalent,
except in N-terminus, to other Mycobacterium DNA GYRASES
SUBUNIT B e.g. T10005 from Mycobacterium leprae (697 aa);
Q9L7L3|GYR_B MYCPA from Mycobacterium avium subsp.
paratuberculosis (677 aa) (has its N-terminus shorter);
P48355|GYR_B MYCSM from Mycobacterium smegmatis (675 aa);
etc. Also highly similar to others e.g. T10969 from
Streptomyces coelicolor (686 aa); P50075|GYR_B_STRSH from
Streptomyces
100.0%; Score 25; DB 1; Length 343050;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 4
MSGY219/c
LOCUS       38721 bp      DNA      linear      BCT 10-DEC-1996
DEFINITION  Mycobacterium tuberculosis sequence from clone y219.
ACCESSION   AD000013
VERSION     AD000013.1  GI:1717736
KEYWORDS    Mycobacterium tuberculosis
SOURCE      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
ORGANISM    Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE   1 (bases 1 to 38721)
            Du, L.,
            Direct Submission
            Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
            Beaver Street, Waltham, MA, USA, 02154 du@eric.com
COMMENT     GSDB:S:1004710
FEATURES    Location/Qualifiers
            source          1..38721
                        /organism="Mycobacterium tuberculosis"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:1773"
                        /clone="y219"

ORIGIN
Query Match      96.0%; Score 24; DB 1; Length 38721;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2  CCGAGGGCCAGAGGCACATAGC 25
      |||||
Db   37274 CCGAGGGCCAGAGGCACATAGC 37251

RESULT 5
AC016557
LOCUS       60740 bp      DNA      linear      HTG 25-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-374J18, *** SEQUENCING IN
ACCESSION   AC016557
VERSION     AC016557.8  GI:15290343
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 60740)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS     DOE Joint Genome Institute
TITLE       Unpublished
JOURNAL
REFERENCE   2 (bases 1 to 60740)
            DOE Joint Genome Institute.
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL
COMMENT     Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Aug 25, 2001 this sequence version replaced gi:13699576.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 8415: contig of 8415 bp in length
            * 8416 8515: gap of unknown length
            * 8516 60740: contig of 52225 bp in length.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            source
            1..168008
            /organism="Homo sapiens"
            /mol_type="genomic DNA"

```

```

Project Information
Center Project Name: 329028
Center clone name: CTC-374J18
-----
Summary Statistics
Consensus quality: 58774 bases at least Q40
Consensus quality: 58860 bases at least Q30
Consensus quality: 58898 bases at least Q20
Estimated insert size: 108000; agarose-fp estimation
Estimated insert size: 60640; sum-of-contigs estimation
Quality coverage: 0.55 in Q20 bases; agarose-fp
estimation
estimation Quality coverage: 0.97 in Q20 bases; sum-of-contigs
estimation.
FEATURES    Location/Qualifiers
            source          1..60740
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="5"
                        /clone="CTC-374J18"

ORIGIN
Query Match      80.8%; Score 20.2; DB 2; Length 60740;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1  CCGGAGGGCCAGAGGCACATAGC 25
      |||||
Db   13110 CCGGATGGCCAGCGGCACAGAGC 13134

RESULT 6
AC008544/c
LOCUS       168008 bp      DNA      linear      PRI 05-APR-2002
DEFINITION  Homo sapiens chromosome 5 clone CTC-500G13, complete sequence.
ACCESSION   AC008544
VERSION     AC008544.7  GI:20043093
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168008)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS     DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL
REFERENCE   2 (bases 1 to 168008)
            DOE Joint Genome Institute.
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL
COMMENT     Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            3 (bases 1 to 168008)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Submitted (05-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            4 (bases 1 to 168008)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Submitted (05-APR-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Apr 5, 2002 this sequence version replaced gi:18497066.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.6% of Sequence;
            Estimated Total Number of Errors is 0.3.
            Location/Qualifiers
            source          1..168008
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"

```

[illegible]

```

REFERENCE
AUTHORS   1 (bases 1 to 183734)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
REFERENCE
AUTHORS   2 (bases 1 to 183734)
TITLE     DOE Joint Genome Institute.
JOURNAL   Direct Submission
REFERENCE
AUTHORS   3 (bases 1 to 183734)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
REFERENCE
AUTHORS   4 (bases 1 to 183734)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
REFERENCE
AUTHORS   5 (bases 1 to 183734)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
COMMENT   On Aug 29, 2003 this sequence version replaced gi:27261496.
          Draft Sequence Produced by DOE Joint Genome Institute
          www.jgi.doe.gov
          Finishing Completed at Stanford Human Genome Center
          www.shgc.stanford.edu
          Quality: Phrap Quality >=40 99.7% of Sequence;
          Estimated Total Number of Errors is 1.2.
FEATURES
source    1. .183734
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /chromosome="5"
          /clone="RP11-443C10"
ORIGIN
Query Match      80.8%; Score 20.2; DB 9; Length 183734;
Best Local Similarity 88.0%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGGGCACATAGC 25
Db 100380 CCGGAGGGCCAGAGGGGCACAGGC 100356

RESULT 9
AC023587/c
LOCUS       AC023587             185329 bp    DNA    linear    HTG 06-AUG-2002
DEFINITION Homo sapiens clone RP11-443C10, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC023587
VERSION     AC023587.3  GI:22123667
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 185329)
AUTHORS     Birren,B., Nusbaum,C. and Lander,B.
TITLE       Homo sapiens, clone RP11-443C10
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 185329)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,P., Boguslavskiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castie,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagsos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Larcocque,K., Lehoczyk,J., Levine,R.,
            Lieh,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
            Direct Submission
            Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 185329)
            Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gind,S., Graham,L., Grand-Pierre,N., Hagsos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Aug 6, 2002 this sequence version replaced gi:7210031.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L6833
            Center clone name: 443_C10
            ----- Summary Statistics
            Sequencing vector: M13; M77815; 42% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 181634 bases at least Q40
            Consensus quality: 182910 bases at least Q30
            Consensus quality: 183679 bases at least Q20
            Insert size: 169000; agarose-fp
            Insert size: 184629; sum-of-contigs
            Quality coverage: 13.2 in Q20 bases; agarose-fp
            Quality coverage: 12.1 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1 5061: contig of 5061 bp in length
            * 5062 5161: gap of 100 bp
            * 5162 6736: contig of 1575 bp in length
            * 6737 6836: gap of 100 bp
            * 6837 9630: contig of 2794 bp in length
            * 9631 9730: gap of 100 bp
            * 9731 13343: contig of 3613 bp in length

```

```

* 13344 13443: gap of 100 bp
* 13444 46306: contig of 32863 bp in length
* 46307 46406: gap of 100 bp
* 46407 93808: contig of 47402 bp in length
* 93809 93908: gap of 100 bp
* 93909 141189: contig of 47281 bp in length
* 141190 141289: gap of 100 bp
* 141290 185329: contig of 44040 bp in length.

FEATURES             Location/Qualifiers
     source            1..185329
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /clone="RP11-443C10"
                        /clone_lib="RPCI-11 Human Male BAC"
     misc_feature       1..5061
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:left"
     misc_feature       5162.._6736
                        /note="assembly_fragment"
     misc_feature       6937.._9830
                        /note="assembly_fragment"
     misc_feature       9731.._13343
                        /note="assembly_fragment"
     misc_feature       13444.._46306
                        /note="assembly_fragment"
     misc_feature       46407.._93808
                        /note="assembly_fragment"
     misc_feature       93909.._141189
                        /note="assembly_fragment"
     misc_feature       141290.._185329
                        /note="assembly_fragment"
                        clone_end:T7
                        vector_side:right"

ORIGIN
Query Match          80.8%; Score 20.2; DB 2; Length 185329;
Best Local Similarity 88.0%; Pred.No.1.le+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CCCGGAGGCCAGAGGCCACATAGC 25
        ||||| ||||| ||||| |||||
Db       101360 CCCGGATGTCGACGGCGCACAGAC 101336

RESULT 10
AC136594
LOCUS      Homo sapiens chromosome 5 clone RP11-1069G21, WORKING DRAFT
DEFINITION linear HTG 06-NOV-2002
ACCESSION AC136594
VERSION    GI:24635911
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 199768)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 5
            Unpublished
REFERENCE  2 (bases 1 to 199768)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (06-NOV-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information

```

SEQUENCE, 24 unordered pieces.

AC136596
 VERSION
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 220632)
 DOE Joint Genome Institute.
 Sequencing of Human Chromosome 5
 Unpublished
 DOE Joint Genome Institute.
 2 (bases 1 to 220632)
 Direct Submission
 Submitted (06-NOV-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

 Project Information
 Center Project Name: 1705510
 Center clone name: RPCI-11_1070F6

 Summary Statistics
 Consensus quality: 199821 bases at least Q40
 Consensus quality: 204729 bases at least Q30
 Consensus quality: 208957 bases at least Q20
 Estimated insert size: 238000; agarose-*fp* estimation
 Estimated insert size: 218332; sum-of-contigs estimation
 Quality coverage: 7.46 in Q20 bases; agarose-*fp* estimation
 Quality coverage: 8.13 in Q20 bases; sum-of-contigs estimation.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1126: contig of 1126 bp in length
 * 1127
 * 1227
 * 1227: contig of 1155 bp in length
 * 2382
 * 2481: gap of unknown length
 * 3622
 * 3721: gap of unknown length
 * 3722
 * 4883: contig of 1162 bp in length
 * 4884
 * 4983: gap of unknown length
 * 4984
 * 6128: contig of 1145 bp in length
 * 6129
 * 6228: gap of unknown length
 * 6229
 * 7648: contig of 1420 bp in length
 * 7649
 * 7748: gap of unknown length
 * 8892: contig of 1144 bp in length
 * 8893
 * 10555: contig of 1563 bp in length
 * 10556
 * 10556: gap of unknown length
 * 10656
 * 12024: contig of 1369 bp in length
 * 12025
 * 12224: gap of unknown length
 * 13593: contig of 1459 bp in length
 * 13594
 * 13694: gap of unknown length
 * 13694
 * 15293: contig of 1600 bp in length
 * 15294
 * 15393: gap of unknown length
 * 15394
 * 16761: contig of 1368 bp in length
 * 16762
 * 16861: gap of unknown length
 * 16862
 * 19310: contig of 2449 bp in length
 * 19311
 * 19410: gap of unknown length
 * 20398: contig of 1588 bp in length
 * 20399
 * 21098: gap of unknown length
 * 21099
 * 22546: contig of 1448 bp in length
 * 22547
 * 22647: gap of unknown length
 * 22647
 * 24398: contig of 1752 bp in length

24399 24498: gap of unknown length
 * 24399
 * 24499
 * 26051
 * 26150: contig of 1552 bp in length
 * 26151
 * 28718: contig of 2568 bp in length
 * 28719
 * 28818: gap of unknown length
 * 28819
 * 33626: contig of 4808 bp in length
 * 33627
 * 33726: gap of unknown length
 * 33727
 * 38907: contig of 5181 bp in length
 * 38908
 * 39007: gap of unknown length
 * 39008
 * 75448: contig of 36441 bp in length
 * 75449
 * 75548: gap of unknown length
 * 115760: contig of 40212 bp in length
 * 115761
 * 115860: gap of unknown length
 * 115861
 * 160813: contig of 44953 bp in length
 * 160814
 * 160913: gap of unknown length
 * 160914
 * 220632: contig of 59719 bp in length.

FEATURES
 source
 1..220632
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-1070F6"
 /clone_lib="RPCI human BAC library 11"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 2; Length 220632;
 Best Local Similarity 88.0%; Pred. No. 1.le+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 CCCGAGGGCCACAGGGCCACATAGC 25
 Db 24048 CCCGATGGCCAGCGGCACAGAGC 24072

RESULT 12
 AC114743
 LOCUS AC114743 104926 bp DNA linear PRI 29-MAY-2002
 DEFINITION Homo sapiens BAC clone RP11-103H17 from 4, complete sequence.
 ACCESSION AC114743
 VERSION AC114743.4 GI:20429584
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 104926)
 Sulston,J.E. and Waterston,R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 99063792
 PUBMED 9847074
 2 (bases 1 to 104926)
 Tomlinson,C., Meyer,R. and Doebber,A.
 The sequence of Homo sapiens BAC clone RP11-103H17
 Unpublished (2001)
 JOURNAL
 3 (bases 1 to 104926)
 Waterston,R.H.
 Direct Submission
 TITLE Direct Submission
 JOURNAL Submitted (11-MAR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 104926)
 Waterston,R.H.
 Direct Submission
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 104926)
 Waterston,R.H.
 Direct Submission
 TITLE Direct Submission
 JOURNAL Submitted (03-MAY-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

MO 63108, USA
6 (Bases 1 to 104926)
Waterston,R.
Direct Submission
Submitted (29-MAY-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 3, 2002 this sequence version replaced gi:20304048.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0103H17

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseegawa K., Woon P.Y., Zhao B., Frengen E., Tatenoh M., Catanese J.J. and de Jong P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-83A24, 2000 bp overlap; the clone sequenced to the right is RP11-308D13, 2000 bp overlap.
Actual start of this clone is at base position 113772 of RP11-83A24; actual end is at base position 61170 of RP11-308D13.

Data from AC112141 was used to finish AC114743.

FEATURES
Source

1. .104926
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-103H17"
/clone_lib="RPCI-11"
3. .140
/rpt_family="L1"
141. .447
/rpt_family="MER2_type"
448. .920
/rpt_family="L1"
1371. 1700
/rpt_family="L1"
1701. .1987
/rpt_family="Alu"
1988. .3303
/rpt_family="L1"

repeat_region 3304. .3594
/rpt_family="Alu"
repeat_region 3619. .4074
/rpt_family="MALR"
repeat_region 4170. .4255
/rpt_family="L1"
repeat_region 4256. .4339
/rpt_family="MER2_type"
4330. .4407
/rpt_family="MER2_type"
4408. .4486
/rpt_family="MER2_type"
4487. .4589
/rpt_family="L1"
5080. .5526
/rpt_family="ERV1"
5527. .5685
/rpt_family="ERV1"
5686. .5979
/rpt_family="Alu"
5980. .7106
/rpt_family="ERV1"
10347. .10654
/rpt_family="Alu"
10659. .11120
/rpt_family="ERV1"
11121. .11405
/rpt_family="Alu"
11477. .11638
/rpt_family="(TG)n"
11791. .11846
/rpt_family="Alu"
11847. .12060
/rpt_family="ERV1"
12061. .12444
/rpt_family="ERV1"
12445. .12464
/rpt_family="(TA)n"
12465. .12504
/rpt_family="ERV1"
13184. .13380
/rpt_family="MER1_type"
13583. .13886
/rpt_family="Alu"
13894. .14049
/rpt_family="Alu"
14052. .14361
/rpt_family="Alu"
15977. .17226
/note="CpG island (%GC=67.1, o/e=0.80, #CpGs=120)"
17300. .17424
/rpt_family="MER1_type"
18508. .18589
/rpt_family="MIR"
18590. .18889
/rpt_family="Achoho"
18690. .18808
/rpt_family="MIR"
19959. .19980
/rpt_family="AT-rich"
20365. .20515
/rpt_family="MER1_type"
20834. .20959
/rpt_family="Alu"
20970. .21795
/rpt_family="L1"
21857. .22138
/rpt_family="Alu"
22317. .22618
/rpt_family="Alu"
22672. .23037
/rpt_family="L2"
23749. .24050
repeat_region

Mon Jun 21 09:02:12 2004

```

repeat_region      /rpt_family="MaLR"
24051..24233
/rpt_family="MaLR"
24504..24604
/rpt_family="Alu"
24742..25040
/rpt_family="Alu"
25407..25493
/rpt_family="ERV1"
25582..25854
/rpt_family="Alu"
25866..26123
/rpt_family="Alu"
26679..26984
/rpt_family="Alu"
27085..28336
/rpt_family="L1"
28403..28615
repeat_region

Query Match      76.8%; Score 19.2; DB 9; Length 104926;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CCGAGGGGCGAGGGGCGACATGC 25
Db      87138 CCGAGGGGCGAGGGGCGACATGC 87161

RESULT 13
AC016802/c
LOCUS
DEFINITION Homo sapiens clone RP11-509, WORKING DRAFT SEQUENCE, 30 unordered
pieces.
ACCESSION AC016802
VERSION AC016802.5 GI:7230035
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146868)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-509
Unpublished
2 (bases 1 to 146868)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferrelita,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6980268.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WISR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2879

```

```

Center clone name: 5_O_9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 129806 bases at least Q40
Consensus quality: 138391 bases at least Q30
Consensus quality: 141610 bases at least Q20
Insert size: 79000; agarose-fp
Quality coverage: 5.8 in Q20 bases; agarose-fp
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1031: contig of 1031 bp in length
* 1032: gap of 100 bp
* 1132: contig of 1063 bp in length
* 2195: gap of 100 bp
* 2295: contig of 1097 bp in length
* 3391: gap of 100 bp
* 3431: gap of 100 bp
* 4032: contig of 541 bp in length
* 4132: gap of 100 bp
* 4133: contig of 1850 bp in length
* 5982: gap of 100 bp
* 6082: contig of 1491 bp in length
* 7573: gap of 100 bp
* 7574: contig of 1453 bp in length
* 9126: gap of 100 bp
* 9227: contig of 1193 bp in length
* 9227: gap of 100 bp
* 10420: contig of 2110 bp in length
* 10520: contig of 2110 bp in length
* 12630: contig of 1363 bp in length
* 12730: gap of 100 bp
* 14092: contig of 1363 bp in length
* 14132: gap of 100 bp
* 14193: contig of 2250 bp in length
* 16443: gap of 100 bp
* 16543: contig of 763 bp in length
* 17305: gap of 100 bp
* 17405: contig of 3171 bp in length
* 20576: gap of 100 bp
* 20577: contig of 2731 bp in length
* 23407: gap of 100 bp
* 23408: contig of 2642 bp in length
* 26150: gap of 100 bp
* 26250: contig of 2988 bp in length
* 29237: gap of 100 bp
* 29338: contig of 3365 bp in length
* 32722: gap of 100 bp
* 32723: contig of 3390 bp in length
* 36212: gap of 100 bp
* 36213: contig of 3810 bp in length
* 40122: gap of 100 bp
* 40222: contig of 3442 bp in length
* 43664: gap of 100 bp
* 43764: contig of 5414 bp in length
* 49178: gap of 100 bp
* 49278: contig of 4197 bp in length
* 53476: gap of 100 bp
* 53575: contig of 6106 bp in length
* 59681: gap of 100 bp
* 59781: contig of 4998 bp in length
* 64779: gap of 100 bp
* 64879: contig of 7167 bp in length
* 72046: gap of 100 bp
* 72146: contig of 9418 bp in length
* 81564: contig of 100 bp
* 81664: gap of 100 bp

```



```
* 81665 94108: contig of 12444 bp in length
* 94109 94208: gap of 100 bp
* 94209 108100: contig of 13892 bp in length
* 108101 108200: gap of 100 bp
* 108201 121190: contig of 12990 bp in length
* 121191 121290: gap of 100 bp
* 121291 146868: contig of 25578 bp in length.
FEATURES
  source
    1..146868
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone_lib="RP11-509"
      /clone_lib="RP11-11 Human Male BAC"
    1..1031
      /note="assembly_fragment"
    1132..2194
      /note="assembly_fragment"
    2295..3391
      /note="assembly_fragment"
    3492..4032
      /note="assembly_fragment"
    clone_end:77
      /note="assembly_fragment"
    vector_side:right
      /note="assembly_fragment"
    4133..5982
      /note="assembly_fragment"
    6083..7573
      /note="assembly_fragment"
    7674..9126
      /note="assembly_fragment"
    9227..10419
      /note="assembly_fragment"
    10520..12629
      /note="assembly_fragment"
    12730..14092
      /note="assembly_fragment"
    14193..16442
      /note="assembly_fragment"
    16543..17305
      /note="assembly_fragment"
    clone_end:SP6
      /note="assembly_fragment"
    vector_side:left
      /note="assembly_fragment"
    17406..20576
      /note="assembly_fragment"
    20677..23407
      /note="assembly_fragment"
    23508..26149
      /note="assembly_fragment"
    26250..29237
      /note="assembly_fragment"
    29338..32722
      /note="assembly_fragment"
    32823..36212
      /note="assembly_fragment"
    36313..40122
      /note="assembly_fragment"
    40223..43664
      /note="assembly_fragment"
    43765..49178
      /note="assembly_fragment"
    49279..53475
      /note="assembly_fragment"
    53576..59681
      /note="assembly_fragment"
    59782..64779
      /note="assembly_fragment"
    64880..72046
      /note="assembly_fragment"
    72147..81564
      /note="assembly_fragment"
    81665..94108
      /note="assembly_fragment"
    94209..108100
```

```
misc_feature
  /note="assembly_fragment"
  108201..121190
misc_feature
  /note="assembly_fragment"
  121291..146868
ORIGIN
Query Match      76.8%; Score 19.2; DB 2; Length 146868;
Best Local Similarity 87.5%; Pred.No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 CCGAGGGCCAGAGGGCCACATAGC 25
    ||||| ||||| ||||| ||||| |||||
Db 133799 CCGAGGGCCAGAGGGCCACAGTGC 133776

RESULT 14
AC112141
LOCUS      166558 bp DNA linear HTG 23-FEB-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-291M3, WORKING DRAFT SEQUENCE,
6 unordered pieces.
ACCESSION  AC112141
VERSION    AC112141.4 GI:18873907
KEYWORDS   HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 166558)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
REFERENCE  2 (bases 1 to 166558)
            Waterston,R.H.
            Direct Submission
            Submitted (19-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE  3 (bases 1 to 166558)
            Waterston,R.H.
            Direct Submission
            Submitted (23-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT    On Feb 23, 2002 this sequence version replaced gi:18860765.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: HNH0291M03
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162870 bases at least Q40
Consensus quality: 163673 bases at least Q30
Consensus quality: 164185 bases at least Q20
Insert size: 172000; agarose-fp
Quality coverage: 7.48 in Q20 bases; agarose-fp
Quality coverage: 7.11 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
```

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1513: contig of 1513 bp in length
 * 1514 1613: gap of unknown length
 * 1614 4371: contig of 2658 bp in length
 * 4372 4371: gap of unknown length
 * 20890: contig of 16519 bp in length
 * 20891 20990: gap of unknown length
 * 20991 64495: contig of 43505 bp in length
 * 64496 64595: gap of unknown length
 * 64596 113460: contig of 48865 bp in length
 * 113461 113560: gap of unknown length
 * 113561 166558: contig of 52998 bp in length.

FEATURES

Location/Qualifiers
 1. .166558
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP11-291M3"
 misc_feature
 1. .1513
 /note="assembly_name:Contig22"
 misc_feature
 1614. .4271
 /note="assembly_name:Contig29"
 misc_feature
 4372. 20890
 /note="assembly_name:Contig30"
 misc_feature
 20991. .64495
 /note="assembly_name:Contig31"
 misc_feature
 64596. .113460
 /note="assembly_name:Contig32"
 misc_feature
 113561. .166558
 /note="assembly_name:Contig33"

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 166558;

Best Local Similarity 87.5%; Pred. No. 3.2e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGGCGCACATAGC 25

Db 124755 CCGAGGGCCAGGCGCACATGTC 124778

RESULT 15

AC084771/c

LOCUS

DEFINITION Homo sapiens chromosome 4 clone RP11-103H17 map 4, WORKING DRAFT

SEQUENCE, 61 unordered pieces.

AC084771

AC084771.2 GI:11693369

VERSION HTGS PHASE1, HTGS_DRAFT.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 169651)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 169651)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Barna,N., Easten,V., Beda,F., Boguslavsky,L.,

Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castie,A.,

Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,

FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,

Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L.,

Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Larocque,K.,

Lamazares,R., Landers,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G.,

Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,

McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,

Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,

O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sougniez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A., and Zody,M.
 Direct Submission
 Submitted (15-NOV-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 13, 2000 this sequence version replaced gi:11178112.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10870

Center clone name: 103_H17

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145723 bases at least Q40

Consensus quality: 156808 bases at least Q30

Consensus quality: 161002 bases at least Q20

Insert size: 186000; agarose-fp

Insert size: 163651; sum-of-contigs

Quality coverage: 2.7 in Q20 bases; agarose-fp

Quality coverage: 3.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 61 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 108: contig of 108 bp in length

* 208: gap of 100 bp

* 209 848: contig of 640 bp in length

* 849 948: gap of 100 bp

* 949 1744: contig of 796 bp in length

* 1745 1844: gap of 100 bp

* 1845 2603: contig of 759 bp in length

* 2604 2703: gap of 100 bp

* 2704 3290: contig of 587 bp in length

* 3291 3391: gap of 100 bp

* 3391 3985: contig of 595 bp in length

* 3986 4085: gap of 100 bp

* 4086 4881: contig of 796 bp in length

* 4882 4981: gap of 100 bp

* 4982 5752: contig of 770 bp in length

* 5752 5852: contig of 100 bp

* 5852 6501: gap of 100 bp

* 6501 7729: contig of 1129 bp in length

* 7730 7829: gap of 100 bp

* 7830 8907: contig of 1078 bp in length

* 8908 9007: gap of 100 bp

* 9008 10139: contig of 1132 bp in length

* 10140 10239: gap of 100 bp

* 10240 11267: contig of 1028 bp in length

* 11268 11367: gap of 100 bp

* 11368 12384: contig of 1017 bp in length

* 12385 12484: gap of 100 bp

* 12485 13466: contig of 982 bp in length

* 13466: contig of 100 bp

* 13467

TITLE

JOURNAL

COMMENT

14768: contig of 1202 bp in length
14769: gap of 100 bp
14770: contig of 18328 bp in length
14771: gap of 100 bp
14772: contig of 1399 bp in length
14773: gap of 100 bp
14774: contig of 1187 bp in length
14775: gap of 100 bp
14776: contig of 1518 bp in length
14777: gap of 100 bp
14778: contig of 1419 bp in length
14779: gap of 100 bp
14780: contig of 1555 bp in length
14781: gap of 100 bp
14782: contig of 1267 bp in length
14783: gap of 100 bp
14784: contig of 1636 bp in length
14785: gap of 100 bp
14786: contig of 1626 bp in length
14787: gap of 100 bp
14788: contig of 1625 bp in length
14789: gap of 100 bp
14790: contig of 1676 bp in length
14791: gap of 100 bp
14792: contig of 2056 bp in length
14793: gap of 100 bp
14794: contig of 1227 bp in length
14795: gap of 100 bp
14796: contig of 1945 bp in length
14797: gap of 100 bp
14798: contig of 1945 bp in length
14799: gap of 100 bp
14800: contig of 2385 bp in length
14801: gap of 100 bp
14802: contig of 2129 bp in length
14803: gap of 100 bp
14804: contig of 1803 bp in length
14805: gap of 100 bp
14806: contig of 2429 bp in length
14807: gap of 100 bp
14808: contig of 2694 bp in length
14809: gap of 100 bp
14810: contig of 2085 bp in length
14811: gap of 100 bp
14812: contig of 3156 bp in length
14813: gap of 100 bp
14814: contig of 2827 bp in length
14815: gap of 100 bp
14816: contig of 3385 bp in length
14817: gap of 100 bp
14818: contig of 2737 bp in length
14819: gap of 100 bp
14820: contig of 2071 bp in length
14821: gap of 100 bp
14822: contig of 3719 bp in length
14823: gap of 100 bp
14824: contig of 3059 bp in length
14825: gap of 100 bp
14826: contig of 3384 bp in length
14827: gap of 100 bp
14828: contig of 3734 bp in length
14829: gap of 100 bp
14830: contig of 3058 bp in length
14831: gap of 100 bp
14832: contig of 3449 bp in length
14833: gap of 100 bp
14834: contig of 3508 bp in length
14835: gap of 100 bp
14836: contig of 3623 bp in length
14837: gap of 100 bp
14838: contig of 3366 bp in length
14839: gap of 100 bp
14840: contig of 3170 bp in length
14841: gap of 100 bp
14842: contig of 4273 bp in length

122976: gap of 100 bp
122977: contig of 4021 bp in length
122978: gap of 100 bp
122979: contig of 3969 bp in length
122980: gap of 100 bp
122981: contig of 6208 bp in length
122982: gap of 100 bp
122983: contig of 4661 bp in length
122984: gap of 100 bp
122985: contig of 4837 bp in length
122986: gap of 100 bp
122987: contig of 6688 bp in length
122988: gap of 100 bp
122989: contig of 6239 bp in length
122990: gap of 100 bp
122991: contig of 8513 bp in length
122992: gap of 100 bp
122993: contig of 739 bp in length

FEATURES
Location/Qualifiers
1..169651
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-103H17"
/clone_lib="RP11-103H17 Human Male BAC"

Query Match 76.8%; Score 19.2; DB 2; Length 169651;
Best Local Similarity 87.5%; Pred. NO. 3.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGCACATAGC 25
Db 157980 CCGAGGGCCAGAGGCACATAGC 157957

RESULT 16

AC110113/c 258723 bp DNA linear HTG 11-OCT-2002
LOCUS Rattus norvegicus clone CH230-303G1, *** SEQUENCING IN PROGRESS
DEFINITION *** 3 unordered pieces.

AC110113
AC110113.4 GI:23605546
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 258723)

REFERENCE

AUTHORS
Muzny, D., Allen, H., Altschuler, S., Amin, A., Anguiano, D., Ayalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenship, K., Blyth, P., Brown, M., Bryant, N., Burch, P., Burrell, K., Caldwell, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Chen, A., Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hernandez, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpthy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J., Lorensheewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelam, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puar, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished
2 (bases 1 to 258723)

Worley, K.C.

Direct Submission

Submitted (10-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 258723)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 9, 2002 this sequence version replaced gi:21738142.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOIN
Center clone name: CH230-303G1
----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 218906 bases at least Q40

Consensus quality: 223719 bases at least Q30

Consensus quality: 226910 bases at least Q20

Estimated insert size: 223284; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 254505: contig of 254505 bp in length

* 254506 254605: gap of unknown length

* 254606 255840: contig of 1235 bp in length

* 255841 255941: gap of unknown length

* 255941 258723: contig of 2783 bp in length.

FEATURES

source

1. 258723

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-303G1"

misc_feature

1. 1320

/note="wgs end_extension"

clone_end:T7"

4253. 6051

/note="wgs end_extension"

clone_end:T7"

7039. 7884

/note="clone_boundary"

clone_end:T7"

site:MboI

end_sequence:RXAKI37TU"

26620. 29134

/note="wgs contig"

237916. 239376

/note="wgs contig"

245651. 248524

/note="clone_boundary"

clone_end:Sp6

site:MboI

end_sequence:RXAKI37TV"

249513. 250744

/note="wgs end_extension"

clone_end:Sp6"

253033. 254505

/note="wgs end_extension"

clone_end:Sp6"

ORIGIN

Query Match

75.2%; Score 18.8; DB 2; Length 258723;

Best Local Similarity 90.9%; Pred.No. 4.5e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACAT 22

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

RESULT 17

MMU93191/c

LOCUS

1276 bp DNA linear ROD 04-NOV-1998

DEFINITION

Mus musculus histone deacetylase-2 (HDAC-2) gene, promoter

ACCESSION

U93191

VERSION

U93191.1

KEYWORDS

GI:3834376

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Zeng, Y., Tang, C.M., Yao, Y.L., Yang, W.M. and Seto, E.

Cloning and characterization of the mouse histone deacetylase-2

gene

J. Biol. Chem. 273 (44), 28921-28930 (1998)

MEDLINE 99003242

```

PUBMED 9786895
REFERENCE 2 (bases 1 to 1276)
AUTHORS Zeng,Y. and Seto,E.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1997) Molecular Oncology, H. Lee Moffitt Cancer
Center, 12902 Magnolia Drive, Tampa, FL 33612, USA
FEATURES
source
1..1276
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129/SvJ"
/db_xref="taxon:10090"
/chromosome="10"
/map="10B1"
/cell_line="ES"
1..1276
/gene="HDAC-2"
1..1276
/gene="HDAC-2"
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 1276;
Best Local Similarity 84.0%; Pred. No. 9.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCACATAGC 25
Db 1128 CCGGAGGGCCGTAGGCACAGC 1104

RESULT 18
AX695575 59817 bp DNA linear PAT 31-MAR-2003
LOCUS AX695575
DEFINITION Sequence 1202 from Patent WO03008583.
ACCESSION AX695575
VERSION AX695575.1 GI:29418727
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Morris,D.W. and Engelhard,E.K.
TITLE Novel compositions and methods for cancer
JOURNAL Patent: WO 03008583-A 1202 30-JAN-2003;
Sagres Discovery (US)
FEATURES
source
1..59817
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 74.4%; Score 18.6; DB 6; Length 59817;
Best Local Similarity 84.0%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCACATAGC 25
Db 8918 CCCGAGGGCCAGCGCGCGTAGC 8942

RESULT 19
AC100484/c
LOCUS AC100484
DEFINITION Mus musculus clone RP23-141N14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC100484
VERSION AC100484.1 GI:17047850
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

1 (bases 1 to 60042)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-141N14
Unpublished
2 (bases 1 to 60042)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhaltier,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,T., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,N., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1544
Center clone name: 141_N_14
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
692: contig of 692 bp in length
693
792: gap of 100 bp
793
1527: contig of 735 bp in length
1528
1627: gap of 100 bp
1628
2335: contig of 708 bp in length
2336
2436: gap of 100 bp
2437
3154: contig of 719 bp in length
3155
3254: gap of 100 bp
3255
3967: contig of 713 bp in length
3968
4067: gap of 100 bp
4068
4767: contig of 700 bp in length
4768
4867: gap of 100 bp
4868
5597: contig of 730 bp in length
5598
5697: gap of 100 bp
5698
6418: contig of 721 bp in length
6419
7097: contig of 579 bp in length
7098
7197: gap of 100 bp
7198
7940: contig of 743 bp in length
7941
8040: gap of 100 bp

```

```
*      8041      8730: contig of 750 bp in length
*      8791      8930: gap of 100 bp
*      8891      9633: contig of 743 bp in length
*      9634      9733: gap of 100 bp
*      9734      10454: contig of 721 bp in length
*     10455      10554: gap of 100 bp
*     10555      11273: contig of 719 bp in length
*     11274      11373: gap of 100 bp
*     11374      12103: contig of 730 bp in length
*     12104      12203: gap of 100 bp
*     12204      12941: contig of 738 bp in length
*     12942      13041: gap of 100 bp
*     13042      13772: contig of 731 bp in length
*     13773      13872: gap of 100 bp
*     13873      14588: contig of 716 bp in length
*     14589      14688: gap of 100 bp
*     14689      15393: contig of 705 bp in length
*     15394      15493: gap of 100 bp
*     15494      16237: contig of 744 bp in length
*     16238      16337: gap of 100 bp
*     16338      17071: contig of 734 bp in length
*     17072      17171: gap of 100 bp
*     17172      17917: contig of 746 bp in length
*     17918      18017: gap of 100 bp
*     18018      18727: contig of 710 bp in length
*     18728      18827: gap of 100 bp
*     18828      19531: contig of 704 bp in length
*     19532      19631: gap of 100 bp
*     19632      20349: contig of 718 bp in length
*     20350      20449: gap of 100 bp
*     20450      21149: contig of 700 bp in length
*     21150      21249: gap of 100 bp
*     21250      21978: contig of 729 bp in length
*     21979      22078: gap of 100 bp
*     22079      22784: contig of 706 bp in length
*     22785      22884: gap of 100 bp
*     22885      23636: contig of 752 bp in length
*     23637      23736: gap of 100 bp
*     23737      24445: contig of 709 bp in length
*     24446      24545: gap of 100 bp
*     24546      25292: contig of 747 bp in length
*     25293      25392: gap of 100 bp
*     25393      26157: contig of 765 bp in length
*     26158      26257: gap of 100 bp
*     26258      27004: contig of 747 bp in length
*     27005      27104: gap of 100 bp
*     27105      27822: contig of 718 bp in length
*     27823      27922: gap of 100 bp
*     27923      28648: contig of 726 bp in length
*     28649      28748: gap of 100 bp
*     28749      29479: contig of 731 bp in length
*     29480      29579: gap of 100 bp
*     29580      30325: contig of 746 bp in length
*     30326      30425: gap of 100 bp
*     30426      31179: contig of 754 bp in length
*     31180      31279: gap of 100 bp
*     31280      32022: contig of 743 bp in length
*     32023      32122: gap of 100 bp
*     32123      32850: contig of 728 bp in length
*     32851      33699: contig of 749 bp in length
*     33700      33799: gap of 100 bp
*     33800      34506: contig of 707 bp in length
*     34507      34606: gap of 100 bp
*     34607      35332: contig of 726 bp in length
*     35333      35432: gap of 100 bp
*     35433      36158: contig of 726 bp in length
*     36159      36258: gap of 100 bp
*     36259      36859: contig of 605 bp in length
*     36860      37122: contig of 100 bp
*     36964      37712: contig of 749 bp in length
*     37713      37812: gap of 100 bp
*     37813      38555: contig of 743 bp in length
```

```
*      38556      38655: gap of 100 bp
*      38656      39377: contig of 722 bp in length
*      39378      39477: gap of 100 bp
*      39478      40181: contig of 714 bp in length
*      40192      40291: gap of 100 bp
*      40292      41020: contig of 729 bp in length
*      41021      41120: gap of 100 bp
*      41121      41810: contig of 690 bp in length
*      41811      41910: gap of 100 bp
*      41911      42649: contig of 739 bp in length
*      42650      42749: gap of 100 bp
*      42750      43498: contig of 749 bp in length
*      43499      43598: gap of 100 bp
*      43599      44305: contig of 707 bp in length
*      44306      44405: gap of 100 bp
*      44406      45156: contig of 751 bp in length
*      45157      45256: gap of 100 bp
*      45257      45973: contig of 717 bp in length
*      45974      46073: gap of 100 bp
*      46074      46782: contig of 709 bp in length
*      46783      46882: gap of 100 bp
*      46883      47588: contig of 706 bp in length
*      47589      47688: gap of 100 bp
*      47689      48406: contig of 718 bp in length
*      48407      48506: gap of 100 bp
*      48507      49227: contig of 721 bp in length
*      49228      49327: gap of 100 bp
*      49328      50057: contig of 730 bp in length
*      50058      50157: gap of 100 bp
*      50158      50910: contig of 753 bp in length
*      50911      51010: gap of 100 bp
*      51011      51756: contig of 746 bp in length
*      51757      51856: gap of 100 bp
*      51857      52566: contig of 710 bp in length
*      52567      52666: gap of 100 bp
*      52667      53402: contig of 736 bp in length
*      53403      53502: gap of 100 bp
*      53503      54250: contig of 746 bp in length
*      54251      54350: gap of 100 bp
*      54351      55050: contig of 700 bp in length
*      55051      55150: gap of 100 bp
*      55151      55872: contig of 722 bp in length
*      55873      55972: gap of 100 bp
*      55973      56691: contig of 719 bp in length
*      56692      56791: gap of 100 bp
```

Query Match 74.4%; Score 18.6; DB 2; Length 60042;
Best Local Similarity 84.0%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 CCGGAGGGCCAGAGGCACATAGC 25
Db      18332 CCCTTAGGCTAGAGACATAGC 18308
```

RESULT 20
HSIMMDL/c

LOCUS
DEFINITION DNA sequence of the human immunoglobulin D segment locus.
ACCESSION X97051.1 GI:1770449

VERSION
KEYWORDS D segment; germ line; immunoglobulin heavy chain; JH segment; VH segment.

SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 56830 to 89967)

JOURNAL
MEDLINE Ravetch, J.V., Siebenlist, U., Korsmeyer, S., Waldmann, T. and Leder, P.
Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes
Cell 27 (3 Pt 2), 583-591 (1981)
93358331

PUBMED
REFERENCE 2 (bases 36339 to 36426; 45954 to 46041; 55207 to 55294; 64614 to 64700)
AUTHORS Siebenlist,U., Ravetch,J.V., Korsmeyer,S., Waldmann,T. and Leder,P.
TITLE Human immunoglobulin D segments encoded in tandem multigenic families
JOURNAL Nature 294 (5842), 631-635 (1981)
MEDLINE 82080701
PUBMED 7312051
REFERENCE 3 (bases 90650 to 91231)
AUTHORS Rabbitts,T.H., Forster,A., Baer,R. and Hamlyn,P.H.
TITLE Transcription enhancer identified near the human C mu immunoglobulin heavy chain gene is unavailable to the translocated c-myc gene in a Burkitt lymphoma
JOURNAL Nature 306 (5945), 806-809 (1983)
MEDLINE 84093586
PUBMED 6419124
REFERENCE 4 (bases 1 to 92588)
AUTHORS Buluwela,L., Albertson,D.G., Sherrington,P., Rabbitts,P.H., Spurr,N. and Rabbitts,T.H.
TITLE The use of chromosomal translocations to study human immunoglobulin gene organization: mapping DH segments within 35 kb of the C mu gene and identification of a new DH locus
JOURNAL EMBO J. 7 (7), 2003-2010 (1988)
MEDLINE 88328988
PUBMED 3138112
REFERENCE 5 (bases 12996 to 13531)
AUTHORS Buluwela,L. and Rabbitts,T.H.
TITLE A VH gene is located within 95 Kb of the human immunoglobulin heavy chain constant region genes
JOURNAL Eur. J. Immunol. 18 (11), 1843-1845 (1988)
MEDLINE 89078478
PUBMED 3144456
REFERENCE 6 (bases 38035 to 52980)
AUTHORS Ichihara,Y., Matsuka,K. and Kurosawa,Y.
TITLE Organization of human immunoglobulin heavy chain diversity gene loci
JOURNAL EMBO J. 7 (13), 4141-4150 (1988)
MEDLINE 89210798
PUBMED 3243276
REFERENCE 7 (bases 87277 to 89821)
AUTHORS Mattila,P.S., Schugk,J., Wu,H. and Makela,O.
TITLE Extensive allelic sequence variation in the J region of the human immunoglobulin heavy chain gene locus
JOURNAL Eur. J. Immunol. 25 (9), 2578-2582 (1995)
MEDLINE 96011870
PUBMED 7589129
REFERENCE 8 (bases 1 to 92588)
AUTHORS Corbett,S.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1996) S. Corbett, MRC, Centre for Protein Eng., Hills Road, Cambridge, CB2 2QH, UK
COMMENT This sequence of the human immunoglobulin D segment locus was determined from three overlapping cosmids; COS23, COS21 and COS24 (described in reference [2] above). All the sequencing and analysis was carried out at the Sanger Centre (Hinxton, Cambs, CB10 1RQ, England) whose financial and technical assistance is gratefully acknowledged.
FEATURES
source location/Qualifiers
1..92588
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q32.33"
/germline
/332..>429
/note="Alu repeat: matches 99. .1 of consensus"
repeat_region
/446..>501
/note="Alu repeat: matches 1. .58 of consensus"
repeat_region
/504..>600
/note="Alu repeat: matches 211. .308 of consensus"
repeat_region
/659..1074

repeat_region /note="2 copies of 208 mer 82 % conserved"
680. .991
/note="3 copies of 104 mer 97 % conserved"
repeat_region 1438. .1771
/note="MUTD element fragment"
repeat_region 2589. .2696
/note="L1 element fragment"
repeat_region 3347. .3636
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 3949. .4002
/note="27 copies of 2 mer 82 % conserved"
repeat_region 3963. .4002
/note="10 copies of 4 mer 90 % conserved"
repeat_region 4557. .4868
/note="3 copies of 104 mer 82 % conserved"
repeat_region 4939. .5227
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 5830. .6197
/note="2 copies of 184 mer 88 % conserved"
repeat_region 9220. .9383
/note="Alu repeat: matches 1. .180 of consensus"
repeat_region 9385. .9453
/note="THE1B element fragment"
repeat_region 9389. .9468
/note="WSTA element fragment"
repeat_region 9592. .10057
/note="3 copies of 22 mer 89 % conserved"
gene 13036. .91231
/gene="IGH@"
CDS join(13036. .13084,13168. .>13481)
/gene="IGH@"
/citation=[5]
/codon_start=1
/product="immunoglobulin heavy chain"
/protein_id="CAA65763.1"
/db_xref="GI:1813977"
/db_xref="REMTREMBL:CAA65763"
/translation="MSVSFLIPVLPLVGLPWGLVSOVLOQSGFGLVKPQSTLSLTCAI
SGDSVSSNSAANNLRQSPSRGLEWLGRTYRSKWINDYAVSVKSRITINPTDSKNQF
SQLNSVTPEDTAVYICAR"
sig_peptide join(13036. .13084,13168. .13178)
/gene="IGH@"
/citation=[5]
mat_peptide 13179. .>13481
/gene="IGH@"
/product="immunoglobulin heavy chain"
/citation=[5]
intron 13085. .13167
/gene="IGH@"
/note="VH6-1 leader intron"
/citation=[5]
V_segment 13179. .13483
/gene="IGH@"
/product="immunoglobulin heavy chain"
/standard_name="VH6-1"
/citation=[5]
repeat_region 13856. .14008
/note="L1 element fragment"
repeat_region 16622. .16883
/note="MSTA element fragment"
repeat_region 16636. .16704
/note="THE1B element fragment"
repeat_region 16757. .16977
/note="THE1B element fragment"
repeat_region 18114. .18345
/note="L1 element fragment"
repeat_region 18445. .18743
/note="L1 element fragment"
repeat_region <18813. .>18987
/note="Alu repeat: matches 308. .133 of consensus"
repeat_region <19004. .>19115
/note="Alu repeat: matches 113. .1 of consensus"
repeat_region 19371. .19470

```

repeat_region /note="L1 element fragment"
20192..20215
repeat_region /note="2 copies of 42 mer 94 & conserved"
20282..20365
repeat_region /note="2 copies of 42 mer 94 & conserved"
21647..21738
repeat_region /note="2 copies of 76 mer 98 & conserved"
21726..21939
repeat_region /note="3 copies of 78 mer 98 & conserved"
22161..22382
repeat_region /note="L1 element fragment"
22752..23135
repeat_region /note="8 copies of 48 mer 89 & conserved"
25123..25301
repeat_region /note="MT1D element fragment"
25695..25760
repeat_region /note="3 copies of 22 mer 83 & conserved"
25697..25760
repeat_region /note="16 copies of 4 mer 83 & conserved"
25700..25763
repeat_region /note="4 copies of 16 mer 83 & conserved"
28283..28318
repeat_region /note="12 copies of 3 mer 83 & conserved"
29990..30265
repeat_region /note="2 copies of 138 mer 81 & conserved"
30572..30617
repeat_region /note="MT1A element fragment"
30683..30712
repeat_region /note="MT1A element fragment"
30801..30872
repeat_region /note="MSTC element fragment"
31700..31825
repeat_region /note="3 copies of 42 mer 99 & conserved"
33714..33730
D_segment /gene="IGH"
/product="immunoglobulin heavy chain"
/note="D1-1"
/citation={8}

```

```

Query Match 74.4%; Score 18.6; DB 9; Length 92588;
Best Local Similarity 84.0%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGAGGCCGACAGGGCACATAGC 25
Db 69447 CACAGAGGCCGACAGGGCACAGC 69423

```

```

RESULT 21
AL929178
LOCUS AL929178 100057 bp DNA linear ROD 18-JAN-2003
DEFINITION Mouse DNA sequence from clone RP23-77M17 on chromosome 2, complete
sequence.
ACCESSION AL929178
VERSION AL929178.7 GI:27803212
KEYWORDS HTG
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Glithero.R.
Direct Submission
Submitted (18-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 18, 2003 this sequence version replaced gi:27764108.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute

```

Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> RP23-77M17 is from the RPI-23 Mouse BAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6.

```

FEATURES             Location/Qualifiers
     source           1..100057
                     /organism="Mus musculus"
                     /mol_type="Genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-77M17"
                     /clone_lib="RPI-23"

```

ORIGIN

```

Query Match 74.4%; Score 18.6; DB 10; Length 100057;
Best Local Similarity 84.0%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGAGGCCGACAGGGCACATAGC 25
Db 22573 CCCTTAGGCTAGAGGACATAGC 22597

```

```

RESULT 22
AL603889/c
LOCUS AL603889 145052 bp DNA linear ROD 05-APR-2002
DEFINITION Mouse DNA sequence from clone RP23-407M20 on chromosome 11,
complete sequence.
ACCESSION AL603889
VERSION AL603889.5 GI:20068524
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Van Hellmond,Z.
Direct Submission
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19031503.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

```


database can be found at
http://www.sanger.ac.uk/projects/c_elegans/wormpep RP23-18012 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6.

FEATURES

source

Location/Qualifiers
 1..151555
 /organism="Mus musculus"
 /mol_type="Genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-18012"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 151555;

Best Local Similarity 84.0%; Pred. No. 5.9e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCACAGAGGCACATAGC 25

Db 140146 CCCTTAGGCTAGAGCAGCATAGC 140170

RESULT 25

AB019441/C

LOCUS

DEFINITION Homo sapiens DNA for immunoglobulin heavy-chain variable region,
 complete sequence, 5 of 5.

ACCESSION

AB019441

VERSION

AB019441.1

KEYWORDS

immunoglobulin heavy chain variable region; VH.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (sites)

Miyata, T., and Honjo, T.

Miyata, T., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H.,

Miyata, T., and Honjo, T.

The complete nucleotide sequence of the human immunoglobulin heavy

chain variable region locus

J. Exp. Med. 188 (11), 2151-2162 (1998)

PUBMED

9841928

REFERENCE

2 (bases 1 to 157090)

Matsuda, F.

Direct Submission

Submitted (05-NOV-1998)

Fumihiko Matsuda, Kyoto University Graduate

School of Medicine, Department of Medical Chemistry, Yoshida

Konocho, Sakyo-ku, Kyoto 606, Japan

(E-mail: fmatsumada@viru1.virus.kyoto-u.ac.jp, Tel: 81-75-753-4371,

Fax: 81-75-753-4388)

Location/Qualifiers

1..157090

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="14"

/map="14q32.33"

/map="14q32.33"

/note="octameric regulatory sequence"

10920..10927

10936..10942

11002..83932

/gene="VH"

/pseudo

Join(11002..11047,11130..11433)

/gene="VH"

/note="4-4"

/pseudo

/codon_start=1

/product="immunoglobulin heavy chain variable region"

/protein_id="BAA75058.1"

exon /db_xref="GI:4512312"
 <11002..11047
 /gene="VH"
 /number=1
 /pseudo
 intron 11048..11129
 /gene="VH"
 exon 11130..11433
 /gene="VH"
 /number=2
 /pseudo
 misc_feature 11434..11440
 /gene="VH"
 /note="7mer RSS: recombination signal sequence cacagtg"
 /pseudo 11464..11472
 /gene="VH"
 /note="9mer RSS: recombination signal sequence acacaaacc"
 /pseudo 17751..17757
 /gene="VH"
 /note="heptameric regulatory sequence"
 /pseudo 17760..17767
 /gene="VH"
 /note="octameric regulatory sequence"
 /pseudo 17776..17782
 /gene="VH"
 /note="1-3"
 /codon_start=1
 /product="immunoglobulin heavy chain variable region"
 /protein_id="BAA75059.1"
 /db_xref="GI:4512313"
 <17860..17905
 /gene="VH"
 /number=1
 /pseudo 17906..17990
 /gene="VH"
 17991..18297
 /gene="VH"
 /number=2
 /pseudo 18298..18304
 /gene="VH"
 /note="7mer RSS: recombination signal sequence cacagtg"
 /pseudo 18328..18336
 /gene="VH"
 /note="9mer RSS: recombination signal sequence tcagaaacc"
 /pseudo <21563..>21864
 /gene="VH"
 /note="3-02.1P"
 /pseudo /codon_start=1
 /product="immunoglobulin heavy chain variable region"
 36321..36327
 /gene="VH"
 /note="heptameric regulatory sequence"
 /pseudo 36330..36337
 /gene="VH"
 /note="octameric regulatory sequence"
 /pseudo 36346..36352
 /gene="VH"
 /note="1-36352"
 /protein_id="BAA75058.1"
 <36435..36480,36566..>36872)
 /gene="VH"

sequence.
AL645923 GI:19572045
VERSION
HTG.
Mus musculus (house mouse)
MUS MUSCULUS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 157361)
Smith, M.
Direct Submission
Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced gi.19335835.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, ENBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-75H1 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk

Location/Qualifiers
1..157361
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/clone="RP23-75H1"
/clone_lib="RPCI-23"

ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 157361;
Best Local Similarity 84.0%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGCACATAGC 25
Db 116583 CCTTAGGGCTAGAGCACATAGC 116607
|||||
RESULT 27
AC092729/c 165116 bp DNA linear MAM 28-OCT-2003
LOCUS
Canis familiaris clone RP81-60B6, complete sequence.
AC092729
AC092729.3 GI:37999253
HTG.
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

/note="1-2"
/pseudo
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="BAA75060.1"
/db_xref="GI:4512314"
<36435..36480
/gene="VH"
/number=1
/pseudo
36481..36565
/gene="VH"
36566..36872
/gene="VH"
/number=2
/pseudo
36873..36879
/gene="VH"
/note="7mer RSS: recombination signal sequence cacagtg"
/pseudo
36903..36911
/gene="VH"
/note="9mer RSS: recombination signal sequence tcagaacc"
/pseudo
<78294..>78475
/gene="VH"
/note="4-01.1p"
/pseudo
/codon_start=1
/product="immunoglobulin heavy chain variable region"
83384..83391
/gene="VH"
/note="octameric regulatory sequence"
/pseudo
83400..83406
/gene="VH"
Join(83485..83533,83617..>83932)
/gene="VH"
/note="6-1"
/pseudo
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="BAA75061.1"
/db_xref="GI:4512315"
<83485..83533
/gene="VH"
/number=1
/pseudo
83534..83616
/gene="VH"
83617..83932
/gene="VH"
/number=2
/pseudo
83933..83939
/note="7mer RSS: recombination signal sequence cacagtg"
83963..83971
/note="9mer RSS: recombination signal sequence acacaaacc"

ORIGIN
Query Match 74.4%; Score 18.6; DB 9; Length 157090;
Best Local Similarity 84.0%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGCACATAGC 25
Db 139905 CACAGAGGGCCAGAGGCACATAGC 139881
|||||
RESULT 26
AL645923 157361 bp DNA linear ROD 29-JUN-2002
LOCUS
Mouse DNA sequence from clone RP23-75H1 on chromosome 13, complete

| | | | | | |
|------------|---|-------------|-----|--------|-----------------|
| LOCUS | CNS01D72 | 169802 bp | DNA | linear | PRI 19-NOV-2001 |
| DEFINITION | Human chromosome 14 DNA sequence BAC R-417P24 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence. | | | | |
| ACCESSION | AL122127 | | | | |
| VERSION | ALI22127.6 | GI:17026193 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| AUTHORS | 1 (bases 1 to 169802) Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brottier, P., Catrolicco, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., Debernardinis, V., Cruaud, C., Gyapay, G., Saurin, W. and Weissenbach, J. | | | | |
| TITLE | Sequencing of the human chromosome 14 | | | | |
| REFERENCE | 2 (bases 1 to 169802) Genoscope. | | | | |
| AUTHORS | Direct Submission | | | | |
| TITLE | Submitted (19-NOV-2001) Genoscope - Centre National de Sequencage : | | | | |
| JOURNAL | BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr | | | | |
| COMMENT | - Web : www.genoscope.cns.fr On Nov 20, 2001 this sequence version replaced gi:14715169. ----- Center: Genoscope / Centre National de Sequencage Center code: GS Web site: http://www.genoscope.cns.fr/ Contact: Segref@genoscope.cns.fr ----- The following BAC sequence is oriented from the T7 to the SP6 end. ----- Assembly program: Phrap; version 2.0 Quality coverage: 7.56x in Q20 bases; sum-of-contigs ----- | | | | |
| FEATURES | Source | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 11439..11536 /note="matching EMBL:G14654 RHdb:RH7974 dbSTS:STS21065 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 Identified using the e-PCR software (G. Schuler)" 39191..39405 /note="matching EMBL:G33053 RHdb:RH67749 dbSTS:STS47676 | | | | |
| ST5 | 1..169802 Location/Qualifiers /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="14" /clone="R-417P24" /clone_lib="RPCI-11" 11427..11641 /note="matching EMBL:G33053 RHdb | | | | |

| | | | | | |
|--|--|-------------|-----|--------|-----------------|
| LOCUS | CNS01D72 | 169802 bp | DNA | linear | PRI 19-NOV-2001 |
| DEFINITION | Human chromosome 14 DNA sequence BAC R-417P24 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence. | | | | |
| ACCESSION | AL122127 | | | | |
| VERSION | ALI22127.6 | GI:17026193 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| AUTHORS | 1 (bases 1 to 169802) Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brottier, P., Catrolicco, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., Debernardinis, V., Cruaud, C., Gyapay, G., Saurin, W. and Weissenbach, J. | | | | |
| TITLE | Sequencing of the human chromosome 14 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 169802) Genoscope. | | | | |
| AUTHORS | Direct Submission | | | | |
| TITLE | Submitted (19-NOV-2001) Genoscope - Centre National de Sequencage : | | | | |
| JOURNAL | BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr) | | | | |
| COMMENT | - Web : www.genoscope.cns.fr On Nov 20, 2001 this sequence version replaced gi:14715169. ----- Genome Center Center: Genoscope / Centre National de Sequencage Center code: GS Web site: http://www.genoscope.cns.fr/ Contact: Segref@genoscope.cns.fr ----- The following BAC sequence is oriented from the T7 to the SP6 end. ----- Summary Statistics Assembly program: Phrap; version 2.0 Quality coverage: 7.56x in Q20 bases; sum-of-contigs ----- | | | | |
| Overall quality chart : | ----- | | | | |
| Range : | bases | | | | |
| 0 : | ----- | | | | |
| 1 - 9 : | ----- | | | | |
| 10 - 19 : | ----- | | | | |
| 20 - 29 : | 11 | | | | |
| 30 - 39 : | 159 | | | | |
| 40 - 49 : | 4325 | | | | |
| 50 - 59 : | 11964 | | | | |
| 60 - 69 : | 13264 | | | | |
| 70 - 79 : | 28113 | | | | |
| 80 - 89 : | 55713 | | | | |
| 90 - 99 : | 56253 | | | | |
| Percentage of bases with a quality value >= 40 : 99 %. | ----- | | | | |
| FEATURES | Location/Qualifiers | | | | |
| Source | 1..169802 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | /chromosome="14" | | | | |
| | /clone="R-417P24" | | | | |
| | /clone_lib="RPCI-11" | | | | |
| | 11427..11641 | | | | |
| | /note="matching EMBL:G33053" | | | | |
| STS | RHdb:RH67749 | | | | |
| | dbSTS:STS47676 | | | | |
| | Identified using the e-PCR software (G. Schuler)" | | | | |
| STS | 11439..11536 | | | | |
| | /note="matching EMBL:G14654" | | | | |
| | RHdb:RH7974 | | | | |
| | dbSTS:STS21065 | | | | |
| | Identified using the e-PCR software (G. Schuler)" | | | | |
| STS | 39191..39405 | | | | |
| | /note="matching EMBL:G33053" | | | | |
| | RHdb:RH67749 | | | | |
| | dbSTS:STS47676 | | | | |
| LOCUS | CNS01D72 | 169802 bp | DNA | linear | PRI 19-NOV-2001 |
| DEFINITION | Human chromosome 14 DNA sequence BAC R-417P24 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence. | | | | |
| ACCESSION | AL122127 | | | | |
| VERSION | ALI22127.6 | GI:17026193 | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| AUTHORS | 1 (bases 1 to 169802) Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brottier, P., Catrolicco, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., Debernardinis, V., Cruaud, C., Gyapay, G., Saurin, W. and Weissenbach, J. | | | | |
| TITLE | Sequencing of the human chromosome 14 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 169802) Genoscope. | | | | |
| AUTHORS | Direct Submission | | | | |
| TITLE | Submitted (19-NOV-2001) Genoscope - Centre National de Sequencage : | | | | |
| JOURNAL | BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr) | | | | |
| COMMENT | - Web : www.genoscope.cns.fr On Nov 20, 2001 this sequence version replaced gi:14715169. ----- Genome Center Center: Genoscope / Centre National de Sequencage Center code: GS Web site: http://www.genoscope.cns.fr/ Contact: Segref@genoscope.cns.fr ----- The following BAC sequence is oriented from the T7 to the SP6 end. ----- Summary Statistics Assembly program: Phrap; version 2.0 Quality coverage: 7.56x in Q20 bases; sum-of-contigs ----- | | | | |
| Overall quality chart : | ----- | | | | |
| Range : | bases | | | | |
| 0 : | ----- | | | | |
| 1 - 9 : | ----- | | | | |
| 10 - 19 : | ----- | | | | |
| 20 - 29 : | 11 | | | | |
| 30 - 39 : | 159 | | | | |
| 40 - 49 : | 4325 | | | | |
| 50 - 59 : | 11964 | | | | |
| 60 - 69 : | 13264 | | | | |
| 70 - 79 : | 28113 | | | | |
| 80 - 89 : | 55713 | | | | |
| 90 - 99 : | 56253 | | | | |
| Percentage of bases with a quality value >= 40 : 99 %. | ----- | | | | |
| FEATURES | Location/Qualifiers | | | | |
| Source | 1..169802 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | /chromosome="14" | | | | |
| | /clone="R-417P24" | | | | |
| | /clone_lib="RPCI-11" | | | | |
| | 11427..11641 | | | | |
| | /note="matching EMBL:G33053" | | | | |
| STS | RHdb:RH67749 | | | | |
| | dbSTS:STS47676 | | | | |
| | Identified using the e-PCR software (G. Schuler)" | | | | |
| STS | 11439..11536 | | | | |
| | /note="matching EMBL:G14654" | | | | |
| | RHdb:RH7974 | | | | |
| | dbSTS:STS21065 | | | | |
| | Identified using the e-PCR software (G. Schuler)" | | | | |
| STS | 39191..39405 | | | | |
| | /note="matching EMBL:G33053" | | | | |
| | RHdb:RH67749 | | | | |
| | dbSTS:STS47676 | | | | |
| LOCUS | CNS01D72 | 169802 bp | DNA | linear | PRI 19-NOV-2001 |
| DEFINITION</ | | | | | |

Search completed: June 20, 2004, 11:44:21
Job time: 704.836 secs

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 143.651 Seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccggaggccagagggcacatagc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N Geneseq 29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002s:*

7: Geneseq2003as:*

8: Geneseq2003bs:*

9: Geneseq2003cs:*

10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | | Match | Length | DB ID | Description |
|------------|-------|---------|--------|-------|-----------|---------------------|-------------|
| | | Match | Length | | | | |
| C 1 | 25 | 100.0 | 1584 | 4 | AAH51956 | AAH51956 Mycobacte | |
| C 2 | 25 | 100.0 | 110000 | 4 | AAI99682 | AAI99682 Mycobacte | |
| C 3 | 25 | 100.0 | 110000 | 4 | AAI99683 | AAI99683 Mycobacte | |
| C 4 | 18.6 | 74.4 | 59817 | 8 | ADA02884 | ADA02884 Human DUS | |
| C 5 | 18.6 | 74.4 | 59817 | 9 | ADB72422 | ADB72422 Human DUS | |
| C 6 | 18.4 | 73.6 | 14646 | 6 | ABK64796 | ABK64796 Human ben | |
| C 7 | 18.2 | 72.8 | 1409 | 4 | ABL05531 | ABL05531 Drosophil | |
| C 8 | 18.2 | 72.8 | 2696 | 4 | ABL05332 | ABL05332 Drosophil | |
| C 9 | 18.2 | 72.8 | 2722 | 4 | ABL05344 | ABL05344 Drosophil | |
| C 10 | 18.2 | 72.8 | 3409 | 4 | ABL05530 | ABL05530 Drosophil | |
| C 11 | 18.2 | 72.8 | 90442 | 8 | ADA03077 | ADA03077 Mouse mCG | |
| C 12 | 18.2 | 72.8 | 90442 | 8 | ADA66361 | ADA66361 Mouse mCG | |
| C 13 | 18.2 | 72.8 | 90442 | 9 | ADB72815 | ADB72815 Mouse mCG | |
| C 14 | 17.6 | 70.4 | 492 | 6 | AAK598106 | AAK598106 Human DNA | |
| C 15 | 17.6 | 70.4 | 637 | 4 | AAK542975 | AAK542975 DNA encod | |
| C 16 | 17.6 | 70.4 | 993 | 6 | AAK598057 | AAK598057 Human DNA | |
| C 17 | 17.6 | 70.4 | 1152 | 9 | ADC86886 | ADC86886 Human GPC | |
| C 18 | 17.6 | 70.4 | 3311 | 9 | ADE28123 | ADE28123 Human NTR | |
| C 19 | 17.2 | 68.8 | 373 | 4 | AAK88002 | AAK88002 Human dig | |
| C 20 | 17.2 | 68.8 | 2000 | 7 | ADA73174 | ADA73174 Rice gene | |
| C 21 | 17.2 | 68.8 | 2266 | 4 | AAK89548 | AAK89548 Human dig | |
| C 22 | 17.2 | 68.8 | 10733 | 6 | ABI99772 | ABI99772 Mouse isc | |
| C 23 | 17.2 | 68.8 | 16038 | 4 | AAK89507 | AAK89507 Human dig | |

| | | | | | | | |
|-----|------|------|------|---|----------|-----------|-----|
| 97 | 15.8 | 67.2 | 1992 | 5 | Aaf28250 | Human | TAN |
| 98 | 16.8 | 67.2 | 2017 | 4 | AAL26420 | Human | bre |
| 99 | 16.8 | 67.2 | 2017 | 4 | Aah72622 | Human | cer |
| 100 | 16.8 | 67.2 | 2017 | 4 | Aah72637 | Human | cer |
| 101 | 16.8 | 67.2 | 2038 | 6 | Abz35495 | Human | gen |
| 102 | 16.8 | 67.2 | 2038 | 7 | Abx62913 | Human | act |
| 103 | 16.8 | 67.2 | 2038 | 7 | Abx62913 | Human | act |
| 104 | 16.8 | 67.2 | 2039 | 8 | ACH00174 | Hypoxia-r | |
| 105 | 16.8 | 67.2 | 2039 | 8 | ACH00174 | Rat parti | |
| 106 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 107 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 108 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 109 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 110 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 111 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 112 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 113 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 114 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 115 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 116 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 117 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 118 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 119 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 120 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 121 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 122 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 123 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 124 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 125 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 126 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 127 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 128 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 129 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 130 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 131 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 132 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 133 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 134 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 135 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 136 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 137 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 138 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 139 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 140 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 141 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 142 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 143 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 144 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 145 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 146 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 147 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 148 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 149 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 150 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 151 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 152 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 153 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 154 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 155 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 156 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 157 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 158 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 159 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 160 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 161 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 162 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 163 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 164 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 165 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 166 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 167 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 168 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |
| 169 | 16.8 | 67.2 | 2039 | 9 | Adc62297 | Hypoxia-r | |

Drug target; growth; organism viability; characterisation; ds.
Mycobacterium tuberculosis.
WO200135317-A1.
17-MAY-2001.
13-NOV-2000; 2000WO-US031152.
12-NOV-1999; 99US-0165086P.
12-NOV-1999; 99US-0165124P.
01-FEB-2000; 2000US-0179531P.
(REGC) UNIV CALIFORNIA.
Eisenberg D, Rotstein SH, Marcotte EM;
WPI; 2001-329193/34.
P-PSDB; AAG81105.
Identifying nucleotide or polypeptide sequence for use as drug target,
involves providing algorithm that analyzes a functional relationship
between nucleotide or polypeptide sequences, and comparing the sequences.
Disclosure; Page 53; 207pp; English.
This invention relates to a method for identifying a nucleotide or
polypeptide sequence that may be a drug target, or essential for growth
or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
tuberculosis proteins which are potential drug targets. The DNA and
protein sequences are used to illustrate the method of the invention. The
method involves providing an unknown nucleotide or polypeptide sequences,
and comparing it to a number of sequences along with at least one
algorithm capable of analysing a functional relationship between
nucleotide and polypeptide sequences. The method is useful for
characterising the function of nucleic acids and polypeptides that may be
useful as a target for a drug or essential for the growth or viability of
an organism
Sequence 1584 BP; 299 A; 539 C; 512 G; 234 T; 0 U; 0 Other;
Query Match 100.0%; Score 25; DB 4; Length 1584;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGGAGGGCCAGAGGCACATAGC 25
Db 1019 CCGGAGGGCCAGAGGCACATAGC 995
RESULT 2
WP Sequence split into 45 fragments LOCUS AA199682 Accession Aa199682
WP Fragment Name Begin End
WP AA199682_00 1 110000
WP AA199682_01 100001 210000
WP AA199682_02 200001 310000
WP AA199682_03 300001 410000
WP AA199682_04 400001 510000
WP AA199682_05 500001 610000
WP AA199682_06 600001 710000
WP AA199682_07 700001 810000
WP AA199682_08 800001 910000
WP AA199682_09 900001 1010000
WP AA199682_10 1000001 1110000
WP AA199682_11 1100001 1210000
WP AA199682_12 1200001 1310000
WP AA199682_13 1300001 1410000
WP AA199682_14 1400001 1510000
WP AA199682_15 1500001 1610000
WP AA199682_16 1600001 1710000

AA13224 Probe #31
Aba54922 Human foe
Aa134577 Probe #32
Aba44482 Human bre
Aba24691 Probe #31
Aak28651 Human bon
Aak03197 Human bra
Aa103130 Probe #31
Aba03159 Human gen
Aba40944 Human foe
Aba97640 Human pan
Aac98944 Human pan
Aba98289 Human pan
Aak76591 Human imm
Aa176591 Human imm
Aba57222 Human col
Acd10994 Human col
Aa117451 Probe #73
Aba62380 Human foe
Aa142359 Probe #11
Aba29716 Probe #81
Aak36600 Human bon
Aak10718 Human bra
Aba36255 Human liv
Aba10598 Human gen
Aba61083 Human foe
Aa140981 Probe #96
Aak35265 Human bon
Aak09376 Human bra
Aba35010 Human liv
Aba09678 Human gen
Aa94702 Rat secre
Aca38747 Prokaryot
Aas83610 DNA encod
Aac06640 Human sec
Aas42946 DNA encod
Aac53800 Arabidops
Aa01750 Human col
Aba81552 Thermus c
Aba05033 Human nov
Aba81554 His tagge
Aba81553 His tagge
Aa105091 Human rep
Aa105219 Human rep
Aa105221 Human rep
Aa105089 Human rep
Aa105217 Human rep
Aa105093 Human rep
Aa198104 Human tes
Aa197984 Human tes
Aa198102 Human tes
Aa197982 Human tes
Aa197986 Human tes
Aa198100 Human tes
Aa88209 Human dig
Aa38324 Human ang
64.0 444 4 AA113224
16 64.0 444 4 ABA54922
16 64.0 444 4 AAI34577
16 64.0 444 4 ABA44482
16 64.0 444 4 ABA24691
16 64.0 444 4 ABA24691
16 64.0 444 4 AAK28651
16 64.0 444 4 AAK03197
16 64.0 444 4 AAI03130
16 64.0 444 5 ABA03159
16 64.0 444 8 ABA40944
16 64.0 466 6 ABA97640
16 64.0 490 3 AAC98944
16 64.0 490 6 ABA98289
16 64.0 495 4 AAK76591
16 64.0 495 4 AAK76591
16 64.0 516 9 ABA57222
16 64.0 528 8 ACD10994
16 64.0 584 8 AAI17451
16 64.0 585 4 ABA62380
16 64.0 585 4 AAI142359
16 64.0 585 4 ABA29716
16 64.0 585 4 AAK36600
16 64.0 585 4 AAK10718
16 64.0 585 4 ABA36255
16 64.0 585 6 ABA10598
16 64.0 590 4 ABA61083
16 64.0 590 4 AAI140981
16 64.0 590 4 AAK35265
16 64.0 590 4 AAK09376
16 64.0 590 4 ABA35010
16 64.0 590 6 ABA09678
16 64.0 607 6 Aa94702
16 64.0 621 7 Aca38747
16 64.0 625 5 Aas83610
16 64.0 650 3 Aac06640
16 64.0 689 4 Aas42946
16 64.0 743 3 Aac53800
16 64.0 822 3 Aa01750
16 64.0 1170 6 ABA81552
16 64.0 1174 7 ABA05033
16 64.0 1188 6 ABA81554
16 64.0 1188 6 ABA81553
16 64.0 1197 4 Aa105091
16 64.0 1197 4 Aa105219
16 64.0 1197 4 Aa105221
16 64.0 1197 4 Aa105089
16 64.0 1197 4 Aa105217
16 64.0 1197 4 Aa105093
16 64.0 1197 4 Aa198104
16 64.0 1197 4 Aa197984
16 64.0 1197 4 Aa198102
16 64.0 1197 4 Aa197982
16 64.0 1197 4 Aa197986
16 64.0 1197 4 Aa198100
16 64.0 1208 4 Aa88209
16 64.0 1278 3 Aa38324
C 243 16
C 244 16
C 245 16
C 246 16
C 247 16
C 248 16
C 249 16
C 250 16
C 251 16
C 252 16
C 253 16
C 254 16
C 255 16
C 256 16
C 257 16
C 258 16
C 259 16
C 260 16
C 261 16
C 262 16
C 263 16
C 264 16
C 265 16
C 266 16
C 267 16
C 268 16
C 269 16
C 270 16
C 271 16
C 272 16
C 273 16
C 274 16
C 275 16
C 276 16
C 277 16
C 278 16
C 279 16
C 280 16
C 281 16
C 282 16
C 283 16
C 284 16
C 285 16
C 286 16
C 287 16
C 288 16
C 289 16
C 290 16
C 291 16
C 292 16
C 293 16
C 294 16
C 295 16
C 296 16
C 297 16
C 298 16
C 299 16
C 300 16
AAH51956/c
ID AAH51956 standard; DNA; 1584 BP.
XX
AC AAH51956;
XX
DT
XX
DE
XX
XX
04-SEP-2001 (first entry)
Mycobacterium tuberculosis potential drug target gene SEQ ID 10.
ALIGNMENTS

WP AA199682_17 1700001 1810000
 WP AA199682_18 1800001 1910000
 WP AA199682_19 1900001 2010000
 WP AA199682_20 2000001 2110000
 WP AA199682_21 2100001 2210000
 WP AA199682_22 2200001 2310000
 WP AA199682_23 2300001 2410000
 WP AA199682_24 2400001 2510000
 WP AA199682_25 2500001 2610000
 WP AA199682_26 2600001 2710000
 WP AA199682_27 2700001 2810000
 WP AA199682_28 2800001 2910000
 WP AA199682_29 2900001 3010000
 WP AA199682_30 3000001 3110000
 WP AA199682_31 3100001 3210000
 WP AA199682_32 3200001 3310000
 WP AA199682_33 3300001 3410000
 WP AA199682_34 3400001 3510000
 WP AA199682_35 3500001 3610000
 WP AA199682_36 3600001 3710000
 WP AA199682_37 3700001 3810000
 WP AA199682_38 3800001 3910000
 WP AA199682_39 3900001 4010000
 WP AA199682_40 4000001 4110000
 WP AA199682_41 4100001 4210000
 WP AA199682_42 4200001 4310000
 WP AA199682_43 4300001 4410000
 WP AA199682_44 4400001 4411529
 ID AA199682 standard; DNA; 4411529 BP.
 XX
 AC AA199682;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-00103840.
 XX
 PR 24-JUN-1998; 98US-00103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 XX
 DR Evaluating strain variation of Mycobacterium tuberculosis, comprises
 XX determining the nucleotide sequence of the strain at positions in the
 XX genome corresponding to positions where M. tuberculosis strains CDC 1551
 XX and H37Rv differ.
 XX
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
 XX
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly

CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1
 XX
 SQ Sequence 4411529 BP; 758565A; 1449993C; 1444602G; 758379T; 0U; 00ther;
 Query Match 100.0%; Score 25; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. NO. 0.59;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CCGGAGGCGCAGAGGCACATAGC 25
 Db 24426 CCGGAGGCGCAGAGGCACATAGC 24450
 RESULT 3
 WP Sequence split into 44 fragments LOCUS AA199683 Accession Aai99683
 WP Fragment Name Begin End
 WP AA199683_00 1 110000
 WP AA199683_01 100001 210000
 WP AA199683_02 200001 310000
 WP AA199683_03 300001 410000
 WP AA199683_04 400001 510000
 WP AA199683_05 500001 610000
 WP AA199683_06 600001 710000
 WP AA199683_07 700001 810000
 WP AA199683_08 800001 910000
 WP AA199683_09 900001 1010000
 WP AA199683_10 1000001 1110000
 WP AA199683_11 1100001 1210000
 WP AA199683_12 1200001 1310000
 WP AA199683_13 1300001 1410000
 WP AA199683_14 1400001 1510000
 WP AA199683_15 1500001 1610000
 WP AA199683_16 1600001 1710000
 WP AA199683_17 1700001 1810000
 WP AA199683_18 1800001 1910000
 WP AA199683_19 1900001 2010000
 WP AA199683_20 2000001 2110000
 WP AA199683_21 2100001 2210000
 WP AA199683_22 2200001 2310000
 WP AA199683_23 2300001 2410000
 WP AA199683_24 2400001 2510000
 WP AA199683_25 2500001 2610000
 WP AA199683_26 2600001 2710000
 WP AA199683_27 2700001 2810000
 WP AA199683_28 2800001 2910000
 WP AA199683_29 2900001 3010000
 WP AA199683_30 3000001 3110000
 WP AA199683_31 3100001 3210000
 WP AA199683_32 3200001 3310000
 WP AA199683_33 3300001 3410000
 WP AA199683_34 3400001 3510000
 WP AA199683_35 3500001 3610000
 WP AA199683_36 3600001 3710000
 WP AA199683_37 3700001 3810000
 WP AA199683_38 3800001 3910000
 WP AA199683_39 3900001 4010000
 WP AA199683_40 4000001 4110000
 WP AA199683_41 4100001 4210000
 WP AA199683_42 4200001 4310000
 WP AA199683_43 4300001 4403765
 ID AA199683 standard; DNA; 4403765 BP.
 AC AA199683;
 DT 15-JAN-2002 (first entry)
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX
 OS Mycobacterium tuberculosis.

| | | |
|----|---|--|
| XX | US6294328-BL. | |
| PN | XX | |
| XX | XX | |
| PD | 25-SEP-2001. | |
| XX | XX | |
| PF | 24-JUN-1998; 98US-00103840. | |
| XX | XX | |
| PR | 24-JUN-1998; 98US-00103840. | |
| XX | (GENO-) INST GENOMIC RES. | |
| PA | XX | |
| PI | Fleischmann RD, White OR, Fraser CM, Venter JC; | |
| PI | WPI; 2001-647261/74. | |
| DR | XX | |
| XX | Evaluating strain variation of Mycobacterium tuberculosis, comprises | |
| PT | determining the nucleotide sequence of the strain at positions in the | |
| PT | genome corresponding to positions where M. tuberculosis strains CDC 1551 | |
| PT | and H37Rv differ. | |
| XX | XX | |
| PS | Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English. | |
| XX | XX | |
| CC | The invention relates to evaluating strain variation within and between | |
| CC | different populations of the tuberculosis bacterial pathogen, | |
| CC | Mycobacterium tuberculosis or related Mycobacterium by determining the | |
| CC | nucleotide sequence of the first strain at positions in the complete | |
| CC | sequence of the genome that correspond to positions that differ in the | |
| CC | nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and | |
| CC | H37Rv (AAI99682). The method is useful for evaluating strain variation of | |
| CC | M. tuberculosis and has valuable application in the fields of | |
| CC | tuberculosis genetics, epidemiology, patient treatment and epidemic | |
| CC | monitoring. Note: The sequence data for this patent did not form part of | |
| CC | the printed specification, but was obtained in electronic format directly | |
| CC | from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1 | |
| XX | XX | |
| SQ | Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890ther; | |
| | Query Match 100.0%; Score 25; DB 4; Length 110000; | |
| | Best Local Similarity 100.0%; Pred.No. 0.59; | |
| | Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0 | |
| Oy | 1 CCCGAGGGCCAGGGCAGCATAGC 25 | |
| | | |
| | 24426 CCCGAGGGCCAGGGCAGCATAGC 24450 | |
| Db | XX | |
| | RESULT 4 | |
| | ADA02684 | |
| ID | ADA02684 standard; DNA; 59817 BP. | |
| XX | ADA02684; | |
| XX | AC | |
| XX | DT | |
| XX | 06-NOV-2003 (first entry) | |
| XX | Human DUSP10 carcinoma associated gene, SEQ ID NO:1202. | |
| XX | XX | |
| KW | Human; carcinoma associated; oncogene; carcinoma; cancer; breast; | |
| KW | prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening; | |
| KW | gene; ds. | |
| OS | Homo sapiens. | |
| XX | XX | |
| PN | WO2003057146-A2. | |
| XX | XX | |
| PD | 17-JUL-2003. | |
| XX | XX | |
| PF | 26-DEC-2002; 2002WO-US041414. | |
| XX | XX | |
| PR | 26-DEC-2001; 2001US-00035832. | |
| XX | (SAGR-) SAGRES DISCOVERY. | |
| PA | XX | |
| XX | Morris DW; | |

```

XX WPI; 2003-587068/55.
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1202; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 59817 BP; 16906 A; 12200 C; 12501 G; 18048 T; 0 U; 162 Other;
Query Match 74.4%; Score 18.6; DB 8; Length 59817;
Best Local Similarity 84.0%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGCGAGGGCGCAGGSCACATAGC 25
DB 8918 CCGCGAGGGCGCAGGCGCGGTAGC 8942
|||||
RESULT 5
ADB72422
ID ADB72422 standard; DNA; 59817 BP.
XX
XX AC ADB72422;
XX
XX AC
XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE Human DUSP10 gene.
XX
XX KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX OS Homo sapiens.
XX
XX WO2003008583-A2.
XX
XX PN
XX
XX PD 30-JAN-2003.
XX
XX PF 26-DEC-2001; 2001WO-US051291.
XX
XX PR 02-MAR-2001; 2001US-00798586.
XX 23-OCT-2001; 2001US-00004113.
XX PR 08-NOV-2001; 2001US-00052482.
XX PR 30-NOV-2001; 2001US-00997722.
XX PR 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX PI Morris DW, Engelhard EK;
XX

```

DR WPI; 2003-239337/23.
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX Claim 1; SEQ ID NO 250; 2304pp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
XX Sequence 59817 BP; 16906 A; 12200 C; 12501 G; 18048 T; 0 U; 162 Other;
SQ
Query Match 74.4%; Score 18.6; DB 9; Length 59817;
Best Local Similarity 84.0%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 CCGGAGGCGGAGGCGGACATAGC 25
DB 8918 CCGGAGGCGGAGGCGGCGGTAGC 8942
RESULT 6
ID ABK64796/c
XX ABK64796 standard; DNA; 14646 BP.
XX
XX AC ABK64796;
XX
XX 18-JUN-2002 (first entry)
DE Human benign prostatic hyperplasia gene #691.
XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX Homo sapiens.
OS
PN WO200212440-A2.
PD 14-FEB-2002.
XX
XX 07-AUG-2001; 2001WO-US024708.
XX
XX 07-AUG-2000; 2000US-0223323P.
PR 05-JUN-2001; 2001US-00873319.
XX
XX (GENE-) GENE LOGIC INC.
PA (NISE) JAPAN TOBACCO INC.
XX
XX Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
PI
PI WPI; 2002-257476/30.
DR
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT detecting expression levels of one or more genes in prostate cells from
PT patient that are differentially regulated compared to normal prostate
PT cells.
XX
XX Disclosure; Page 373-377; 444pp; English.
XX
XX The invention relates to a method of diagnosing (I) the onset or
XX progression of benign prostatic hyperplasia (BPH), or screening (II) for
XX or identifying an agent that modulates the onset or progression of BPH.
XX The method is based on changes in gene expression in BPH tissue isolated
XX from patients exhibiting different clinical states of prostate
XX hyperplasia as compared to normal prostate tissue. (I) comprises
XX detecting the expression levels of one or more genes in prostate cells
XX from the subject that are differentially regulated compared to normal
XX prostate cells. (II) comprises preparing a first gene expression profile
XX of BPH cells or BPH-like cell population, exposing the cells to the

CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles. (I)
CC is useful for diagnosing the onset or progression of BPH. (II) is useful
CC for identifying an agent that modulates the onset or progression of BPH.
CC The methods are useful to present information identifying the expression
CC level in a tissue or cells, by comparing the expression level of genes
CC given in the specification in the tissue or cells to the level of
CC expression of gene in the database, and displaying the expression levels
CC of at least one gene in the tissue or cell sample compared to the
CC expression level in BPH. Agents using (II) are useful for treating BPH or
CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic
CC hyperplasia gene sequences of the invention
XX
XX Sequence 14646 BP; 4552 A; 2957 C; 2741 G; 4395 T; 0 U; 1 Other;
SQ
Query Match 73.6%; Score 18.4; DB 6; Length 14646;
Best Local Similarity 95.0%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 5 GAGGGCCAGAGGCGCACATAG 24
DB 2885 GAGGGCCAGAGGCGCACATAG 2866
RESULT 7
ABL05531/c
ID ABL05531 standard; cDNA; 1409 BP.
XX
XX AC ABL05531;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 11075.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUN-2000; 2000US-00614150.
XX
XX (PEKE) PB CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
PI WPI; 2001-656860/75.
DR P-PSDB; ABB61428.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 11075; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1409 BP; 338 A; 375 C; 385 G; 311 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 4; Length 1409;
Best Local Similarity 87.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CGGAGGCCAGAGGCACATAGC 25
|||||
Db 566 CGGATGCCAGAGGCACATTGC 544

RESULT 8
ABL05332/c
ID ABL05332 standard; cDNA; 2696 BP.

XX ABL05332;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 10478.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacological; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB61229.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 10478; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

QY 3 CGGAGGCCAGAGGCACATAGC 25
|||||
Db 151 CGGATGCCAGAGGCACATTGC 129

RESULT 9
ABL05344/c
ID ABL05344 standard; cDNA; 2722 BP.

XX
XX

ABL05344;
26-MAR-2002 (first entry)
Drosophila melanogaster expressed polynucleotide SEQ ID NO 10514.
Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical; gene; ss.
Drosophila melanogaster.
WO200171042-A2.
27-SEP-2001.
23-MAR-2001; 2001WO-US009231.
23-MAR-2000; 2000US-0191637P.
11-JUL-2000; 2000US-00614150.
(PEKE) PE CORP NY.
Venter JC, Adams M, Li PWD, Myers EW;
WPI; 2001-656860/75.
P-PSDB; ABB61241.
New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions.
Claim 1; SEQ ID NO 10514; 21pp + Sequence Listing; English.
The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences

QY 3 CGGAGGCCAGAGGCACATAGC 25
|||||
Db 2421 CGGATGCCAGAGGCACATTGC 2399

RESULT 10
ABL05530/c
ID ABL05530 standard; cDNA; 3409 BP.

XX ABL05530;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 11072.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacological; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX
XX 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR WPI: 2001-656960/75.
 XX DR P-PSDB; ABB61427.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX PS Claim 1; SEQ ID NO 11072; 21bp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 3409 BP; 822 A; 842 C; 862 G; 883 T; 0 U; 0 Other;
 Query Match 72.8%; Score 18.2; DB 4; Length 3409;
 Best Local Similarity 87.0%; Pred. No. 3.5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 CGGAGGGCCAGAGGCACATAGC 25
 DB 1566 CGGATGGCCAGAGACACATTGC 1544
 RESULT 11
 ADA03077
 ID ADA03077 standard; DNA; 90442 BP.
 XX AC ADA03077;
 XX DT 06-NOV-2003 (first entry)
 XX DE Mouse MCG2257 carcinoma associated gene, SEQ ID NO:1595.
 XX KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
 KW gene; ds.
 XX OS Mus sp.
 XX PN WO2003057146-A2.
 XX PD 17-JUL-2003.
 XX PF 26-DEC-2002; 2002WO-US041414.
 XX PR 26-DEC-2001; 2001US-00035832.
 XX PA (SAGR-) SAGRES DISCOVERY.
 XX PI Morris DW;
 XX DR WPI: 2003-587068/55.
 XX PT New recombinant nucleic acid encoding carcinoma associated protein,
 PT useful for preparing compositions for treating carcinomas.

XX Claim 1; SEQ ID NO 1595; 245pp; English.
 XX PS The invention relates to recombinant carcinoma associated (CA) nucleic
 CC acid sequences from mouse and human (ADA01492-ADA03094), and to
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The
 CC invention also encompasses expression vectors and host cells comprising a
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
 CC binds to the protein, and a biochip comprising CA nucleic acid or
 CC fragments thereof. The sequences of the invention were identified using
 CC oncogenic retroviruses, which insert into the genome of the host organism
 CC at random. Many of these do not carry transduced host oncogenes or
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
 CC direct consequence of the effects of proviral integration into host
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
 CC leukaemia) or a propensity to carcinoma by determination of the sequence
 CC of a CA gene, or by determination of CA gene expression in particular
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as
 CC therapeutic agents and in screening and evaluating drug candidates. The
 CC present sequence represents a specifically claimed murine CA nucleic acid
 CC sequence of the invention. Note: the complete sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;
 Query Match 72.8%; Score 18.2; DB 8; Length 90442;
 Best Local Similarity 87.0%; Pred. No. 4.2e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CCGGAGGGCCAGAGGCACATAG 24
 DB 27716 CCAGAGGGACAGAGGCACATAG 27738
 RESULT 12
 ADA66361
 ID ADA66361 standard; DNA; 90442 BP.
 XX AC ADA66361;
 XX DT 20-NOV-2003 (first entry)
 XX DE Mouse MCG2257 gene genomic DNA sequence.
 XX KW carcinoma-associated gene; CA gene; Rorc gene; MCG15938 gene; BAT1 gene;
 KW Igkap1 gene; IGAPI gene; Zfp29 gene; hCG27579 gene; Kcnj9 gene;
 KW Kcnj9 gene; Pp3cc gene; Pp3cc gene; MCG9110 gene; hCG27579 gene;
 KW cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;
 KW carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;
 KW vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;
 KW prostate cancer; DNA vaccine; animal model; mouse; murine; ds; MCG2257.
 XX OS Mus sp.
 XX PN WO2003053224-A2.
 XX PD 03-JUL-2003.
 XX PF 20-DEC-2002; 2002WO-US041776.
 XX PR 20-DEC-2001; 2001US-00034650.
 XX PA (SAGR-) SAGRES DISCOVERY.
 XX PI Morris DW, Engelhard EK;
 XX DR WPI: 2003-569168/53.
 XX PT Novel recombinant carcinoma-associated nucleic acid, useful for
 PT evaluating the effect of a candidate carcinoma drug, and for diagnosing

PT carcinoma.

XX Claim 1; Page 161-174; 229pp; English.

XX This invention relates to a novel recombinant carcinoma-associated (CA)

CC nucleic acid comprising a fully defined genomic, mRNA or coding sequences

CC of mouse Rorc gene or human RORC gene, mouse MCG15938 or human gene BAT1,

CC mouse Igkap1 gene or human Iqgap1 gene, mouse Zfp29 gene or human

CC hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Pp3cc gene or

CC human Pp3CC gene, mouse MCG9110 gene or human hCG27579 gene, as given in

CC the specification. CA genes are genes which are preferably expressed in

CC cancer cells, preferably lymphatic, breast, prostate or epithelial cells.

CC A compound which modifies the expression of the CA genes or bind to

CC carcinoma-associated proteins (CAP) may have cytostatic activity and the

CC sequences of the invention may enable the use of gene therapy or a

CC development of an anticancer vaccine. Therefore the invention may be

CC useful for diagnosis and treatment of carcinomas, especially lymphoma

CC carcinoma, breast cancer and prostate cancer. The CA genes may also be

CC useful as DNA vaccines and for generating animal models of carcinomas.

CC The present sequence is that of the mouse MCG2257 gene genomic DNA

CC sequence of the invention.

XX

SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 8; Length 90442;

Best Local Similarity 87.0%; Pred. No. 4.2e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCAGATAG 24

Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 13

ID ADB72815

AD BDB72815 standard; DNA; 90442 BP.

XX

AC ADB72815;

XX

XX 04-DEC-2003 (first entry)

DT

DE Mouse MCG2257 gene.

XX

XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;

KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX

XX Mus sp.

OS

XX WO2003008583-A2.

XX

XX 30-JAN-2003.

PD

XX 26-DEC-2001; 2001WO-US051291.

PF

XX 02-MAR-2001; 2001US-00798586.

PR

XX 23-OCT-2001; 2001US-00004113.

PR

XX 08-NOV-2001; 2001US-00052482.

PR

XX 30-NOV-2001; 2001US-00997722.

PR

XX 20-DEC-2001; 2001US-00034650.

XX

PA (SAGR-) SAGRES DISCOVERY.

XX

XX Morris DW, Engelhard EK;

PI

XX WPI; 2003-239337/23.

DR

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,

PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX

XX Claim 1; SEQ ID NO 643; 2304pp; English.

PS

XX The invention relates to a novel recombinant nucleic acid comprising a

CC nucleotide sequence selected from any of the 660 sequences fully defined

CC in the specification. A polynucleotide of the invention has cytostatic

CC activity, and may have a use in gene therapy, or in a vaccine. The

CC recombinant nucleic acids and polypeptides are useful for treating

CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

CC sarcomas. The present sequence represents a mouse gene of the invention.

XX

SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 9; Length 90442;

Best Local Similarity 87.0%; Pred. No. 4.2e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCAGATAG 24

Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 14

ID AAS98106

AA AAS98106 standard; DNA; 492 BP.

XX

AC AAS98106;

XX

XX 12-MAR-2002 (first entry)

DT

XX Human DNA for potential G protein-coupled receptor #63.

DE

XX Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;

KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;

KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;

KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;

KW depression; epilepsy; macular degeneration; lymphoma; melanoma;

KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;

KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;

KW tuberculosis; cognition disorder; memory disorder; anorexia;

KW hormonal release disorder; cardiovascular activity disorder;

KW pain perception disorder; obesity; diabetes; obesity; diabetes;

KW hyperlipidaemia; stroke; gene therapy.

XX

OS Homo sapiens.

XX

XX WO200185791-A1.

PN

XX 15-NOV-2001.

PD

XX 11-MAY-2001; 2001WO-US015332.

PF

XX 11-MAY-2000; 2000US-0203217P.

PR

XX 18-MAY-2000; 2000US-0205945P.

PR

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

PA

XX Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;

PI

XX WPI; 2002-066595/09.

DR

XX Novel G protein-coupled receptor polypeptides including galanin receptor

PT polypeptides useful for identifying modulators that are useful for

PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,

PT stroke.

XX

XX Claim 2; Page 81; 144pp; English.

PS

XX The invention relates to an isolated polypeptide encoded by a nucleic

CC acid molecule that is at least 80% identical to the G protein-coupled

CC (GPCR) polynucleotides included in the specification. Also included are

CC probes based on the GPCR sequences (including antisense probes), a host

CC cell comprising an expression vector comprising the GPCR sequence,

CC antibodies raised against the polypeptides, and methods of identifying

CC modulators of the polypeptides. The polypeptides are useful for

CC identifying modulator compounds which function as modulators, activators,

CC repressors, agonists or antagonists of the novel GPCR polypeptides

CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as

described above can be used to detect the presence of the polypeptides and nucleic acids and are used to diagnose a variety of diseases or disorders in which GPCRs are involved e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease, depression, epilepsy, macular degeneration, lymphoma, melanoma, multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other diseases listed in the specification. The probes and antibodies are also useful for diagnosing cognition and memory disorders, anorexia, hormonal release disorders, cardiovascular activity disorders, pain perception disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds that decrease or increase the expression of galanin receptor (GAL4) can be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is useful for treating the above mentioned disorders by gene therapy techniques. The present sequence is a novel GPCR polynucleotide of the invention

Sequence 492 BP; 137 A; 141 C; 128 G; 86 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 492;
Best Local Similarity 83.3%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGGCACATAG 24
DB 109 CCAGTGGGACAGAGGGGCACAG 132

RESULT 15

AAS42975

ID AAS42975 standard; cDNA; 637 BP.

XX AAS42975;

XX 18-DEC-2001 (first entry)

DE DNA encoding G protein-coupled receptor, nGPCR-2079.

XX Human; mental disorder; thyroid disease; renal failure; anorexia;
KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
KW Huntington's disease; Alzheimer's disease; viral infection;
KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
KW nGPCR; ss.

OS Homo sapiens.

XX WO200162924-A2.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-US005989.

XX 24-FEB-2000; 2000US-0184602P.

PR 24-FEB-2000; 2000US-0184604P.

PR 24-FEB-2000; 2000US-0184606P.

PR 24-FEB-2000; 2000US-0184689P.

PR 24-FEB-2000; 2000US-0184690P.

PR 24-FEB-2000; 2000US-0184710P.

PR 24-FEB-2000; 2000US-0184712P.

PR 24-FEB-2000; 2000US-0184715P.

PR 24-FEB-2000; 2000US-0184716P.

PR 24-FEB-2000; 2000US-0184725P.

PR 24-FEB-2000; 2000US-0184822P.

XX (PHAA) PHARMACIA & UPJOHN CO.

PA Vogeli G, Wood LS, Parodi LA, Lind P;

XX WPI; 2002-066595/09.

DR WPI: 2001-570632/64.

XX P-PSDB; AAU25665.

XX Novel nucleic acid and encoded nGPCR-x, used to screen for compounds for use in the treatment of mental disorders, such as Alzheimer's disease, or Parkinson's disease.

XX Claim 3; Page 75; 263pp; English.

XX The invention relates to novel isolated human G protein-coupled receptors (nGPCR-x). The nGPCR-x can be used for screening compounds which can be used to treat mental disorders, thyroid disease, renal failure, inflammatory conditions such as Crohn's disease, rheumatoid arthritis, autoimmune disorders, schizophrenia, migraine, stroke, dementia, depression, Parkinson's disease, Alzheimer's disease, and Huntington's disease. They may also be used for treating viral infections such as human immunodeficiency virus (HIV), type 2 diabetes, obesity, anorexia, hypotension, hypertension, thrombosis, myocardial infarction, atherosclerosis, cancer, and sexual dysfunction. AAS42927-AAS43036 represent the coding sequences of novel human G protein-coupled receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the invention

XX Sequence 637 BP; 177 A; 175 C; 164 G; 121 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 637;
Best Local Similarity 83.3%; Pred. No. 5.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCAGAGGGGCACATAG 24

DB 182 CCAGTGGGACAGAGGGGCACAG 205

RESULT 16

AAS98057/c

ID AAS98057 standard; DNA; 993 BP.

XX AAS98057;

XX 12-MAR-2002 (first entry)

DE Human DNA for potential G protein-coupled receptor #15.

XX Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW tuberculosis; cognition disorder; memory disorder; anorexia;
KW hormonal release disorder; cardiovascular activity disorder;
KW pain perception disorder; obesity; diabetes; obesity; diabetes;
KW hyperlipidaemia; stroke; gene therapy.

OS Homo sapiens.

XX WO200185791-A1.

XX 15-NOV-2001.

XX 11-MAY-2001; 2001WO-US015332.

XX 11-MAY-2000; 2000US-0203217P.

PR 18-MAY-2000; 2000US-0209945P.

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

XX Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;

XX WPI; 2002-066595/09.

PT Novel G protein-coupled receptor polypeptides including galanin receptor
PT polypeptides useful for identifying modulators that are useful for
PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT stroke.
XX
PS Disclosure; Page 121-122; 144pp; English.
XX
CC The invention relates to an isolated polypeptide encoded by a nucleic
CC acid molecule that is at least 80% identical to the G protein-coupled
CC (GPCR) polynucleotides included in the specification. Also included are
CC probes based on the GPCR sequences (including antisense probes), a host
CC cell comprising an expression vector comprising the GPCR sequence,
CC antibodies raised against the polypeptides, and methods of identifying
CC modulators of the polypeptides. The polypeptides are useful for
CC identifying modulator compounds which function as modulators, activators,
CC repressors, agonists or antagonists of the novel GPCR polypeptides
CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC described above can be used to detect the presence of the polypeptides
CC and nucleic acids and are used to diagnose a variety of diseases or
CC disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC macular degeneration, lymphoma, melanoma, multiple sclerosis,
CC osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC diseases listed in the specification. The probes and antibodies are also
CC useful for diagnosing cognition and memory disorders, pain perception
CC release disorders, cardiovascular activity disorders, pain perception
CC disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC that decrease or increase the expression of galanin receptor (GAL4) can
CC be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
CC nucleic acid is useful for treating the above mentioned disorders by gene
CC therapy techniques. The present sequence is a novel GPCR polynucleotide
CC of the invention
XX
SQ Sequence 993 BP; 213 A; 261 C; 248 G; 271 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 6; Length 993;
Best Local Similarity 83.3%; Pred. No. 5.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCACATAG 24
Db 648 CCAGGTGGGACAGAGGGCACAGAG 625
RESULT 17
ADC86886/c
ID ADC86886 standard; DNA; 1152 BP.
XX
AC ADC86886;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:1339.
XX
KW ds; gene; human; GPCR;
XX guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX Homo sapiens.
XX
OS EP1270724-A2.
XX
FN 02-JAN-2003.
XX
PD 18-JUN-2002; 2002EP-00013517.
XX
PF 18-JUN-2001; 2001JP-00246789.
XX
PR (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PT

PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX WPI; 2003-315783/31.
DR P-PSDB; ADC86887.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1339; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1152 BP; 221 A; 303 C; 307 G; 321 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 9; Length 1152;
Best Local Similarity 83.3%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCACATAG 24
Db 554 CCAGGTGGGACAGAGGGCACAGAG 531
RESULT 18
ADE28123
ID ADE28123 standard; cDNA; 3311 BP.
XX
AC ADE28123;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human NTRAN cDNA - SEQ ID 28.
XX
KW human; neurotransmission-associated protein; NTRAN; cytostatic;
XX immunomodulator; immune disorder; cancer; gene therapy; ss; gene.
XX Homo sapiens.
XX
OS WO2003051902-A1.
XX
PN 26-JUN-2003.
XX
PD 12-DEC-2002; 2002WO-US040059.
XX
PF 14-DEC-2001; 2001US-0340798P.
XX
PR 18-MAR-2002; 2002US-0365645P.
XX
PR 25-MAR-2002; 2002US-0367662P.
XX
PR 10-MAY-2002; 2002US-0379887P.
XX
PR 31-MAY-2002; 2002US-0384639P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Baughn MR, Bhatia U, Blake JJ, Burrill JD, Elliott VS;
XX Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AE, Griffin JA;
XX Hafalia AIA, Ho A, Jackson AA, Jiang X, Kable AE, Kearney L;
XX Khare R, Lee S, Lu DAM, Marquis JP, Lehn-Wasson PW;
XX Ramkumar J, Richardson TW, Sprague WW, Tran UK, Chawla NK;
XX Warren BA, Yue H, Zheng W;
XX WPI; 2003-514037/48.
XX
DR P-PSDB; ADE28101.
XX
XX New human neurotransmission-associated proteins (NTRAN) polypeptide,
XX PT useful for preparing a composition for treating a disease associated with
XX decreased expression or overexpression of NTRAN e.g., cancer.
XX

XX Claim 5; SEQ ID NO 28; 261pp; English.
XX The invention relates to a novel isolated human neurotrophin-
CC associated proteins (NTRN) polypeptide. The polypeptide of the invention
CC demonstrates cytotatic and immunomodulator activities and may be useful
CC for preparing a composition for diagnosing or treating a disease or
CC condition associated with decreased expression or overexpression of
CC functional NTRN including immune disorders or cancer, as well as during
CC gene therapy procedures. The current sequence is that of the human NTRN
CC cDNA of the invention.
XX
SQ Sequence 3311 BP; 676 A; 1038 C; 1058 G; 539 T; 0 U; 0 Other;
Query Match 70.4%; Score 17.6; DB 9; Length 3311;
Best Local Similarity 83.3%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CCGAGGGCCAGGACATAGC 25
Db 81 CCGAGGGCCAGGACATAGC 104
RESULT 19
AAK88002/c
ID AAK88002 standard; cDNA; 373 BP.
XX AC AAK88002;
XX
XX 05-NOV-2001 (first entry)
DE Human digestive system antigen coding sequence SEQ ID NO: 318.
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ss.
XX Homo sapiens.
OS
PN WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001324.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180629P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217498P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 03-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

XX 20-NOV-2003 (first entry)
XX Rice gene, SEQ ID 6500.
XX Plant; bacterial infection; fungal infection; viral infection; rice;
XX gene, ds.
XX Oryza sativa.
XX WO2003000898-A1.
XX 03-JAN-2003.
XX 22-JUN-2001; 2001WO-IB001105.
XX 22-JUN-2001; 2001WO-IB001105.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX Identifying at least one gene involved in plant resistance or response to
XX pathogenic infection for conferring resistance or tolerance to a plant to
XX bacterial, fungal or viral infection by determining or detecting plant
XX gene expression.
XX Claim 27; SEQ ID NO 6500; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
XX involved in plant resistance or response to pathogenic infection. M1
XX comprises identifying a gene whose expression is significantly altered in
XX the incompatible interaction of plant gene expression relative to
XX expression of the gene in an uninfected plant, in a mutant plant that
XX does not express a gene associated with response to pathogenic infection,
XX or in a corresponding incompatible or compatible interaction. (M1) is
XX useful for conferring resistance to resistance or tolerance to a plant to
XX bacterial, fungal or viral infection. The present sequence was used to
XX illustrate the invention.
XX Sequence 2000 BP; 656 A; 391 C; 358 G; 593 T; 0 U; 2 Other;
Query Match 68.8%; Score 17.2; DB 7; Length 2000;
Best Local Similarity 86.4%; Pred. No. 8.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CCGAGGGCCGAGGGGACACATA 23
Db 1819 CCGAGGGCCGAGGGGACACATA 1840
RESULT 21
AAK89548/c
ID AAK89548 standard; DNA; 2266 BP.
XX AAK89548;
XX 05-NOV-2001 (first entry)
XX Human digestive system antigen genomic sequence SEQ ID NO: 3124.
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX digestive system disorder; Meckel's diverticulum; ds.
XX Homo sapiens.
XX WO200155314-A2.
XX 02-AUG-2001.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249285P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX P-PSDB; AAM92229.
XX polynucleotides encoding digestive system antigens, useful for
XX diagnosing, treating, preventing and/or prognosing disorders of the
XX digestive system, particularly cancer and cancer metastases.
XX Claim 1; SEQ ID NO 318; 986pp; English.
XX The present invention provides the protein and coding sequences of a
XX number of human digestive system antigens. These can be used in the
XX diagnosis, treatment and prevention of digestive system disorders,
XX including cancer, Meckel's diverticulum, bacterial or parasitic
XX infections, appendicitis, Hirschsprung's disease, chronic colitis or
XX ulcerative colitis. The present sequence is a cDNA encoding a digestive
XX system antigen of the invention
XX Sequence 373 BP; 71 A; 107 C; 101 G; 91 T; 0 U; 3 Other;
Query Match 68.8%; Score 17.2; DB 4; Length 373;
Best Local Similarity 86.4%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 GGAGGGCCGAGGGGACATAGC 25
Db 297 GGAGGGCCGAGGGGACATAGC 276
RESULT 20
ADA73174
ID ADA73174 standard; DNA; 2000 BP.
XX ADA73174;
XX

XX
PF 17-JAN-2001; 2001WC-US001324.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189974P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220363P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0228272P.
PR 22-AUG-2000; 2000US-0228681P.
PR 22-AUG-2000; 2000US-0228688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0228343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231246P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244517P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2000US-0254097P.
PR 05-JAN-2001; 2000US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-502630/55.

XX

PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
XX
PS Disclosure; SEQ ID NO 3124; 986bp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
XX Sequence 2266 BP; 486 A; 594 C; 578 G; 608 T; 0 U; 0 Other;
SQ Query Match 68.8%; Score 17.2; DB 4; Length 2266;
Best Local Similarity 86.4%; Pred. No. 9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 GGAGGGCCAGAGGCACATAGC 25
DB 712 GGAGGGCCAGAGGGGACACAGC 691
RESULT 22
ABI99772/C
ID ABI99772 standard; cDNA; 10733 BP.
XX
XX AC ABI99772;
XX
XX 07-MAR-2002 (first entry)
XX
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:859.
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX Mus musculus.
XX
XX WO200198188-A2.
XX
XX 22-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-JP004192.
XX
XX 18-MAY-2000; 2000JP-00145977.
XX
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
XX WPI; 2002-034733/04.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX expression levels of particular genes defined in the specification or by
XX determining the expression profile of a gene group comprising these
XX genes.
XX
XX Claim 2; Page 2137-2143; 2690pp; English.
XX
XX The present invention describes a method for examining ischaemic
XX conditions, comprising measuring the expression levels of particular
XX genes (i) in a test sample or determining the expression profile of a
XX gene group in the sample comprising genes selected from (i). The method
XX is useful for examining the ischaemic condition (e.g. compressive
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
XX expression levels of particular genes (ABI99202 to ABI99912, encoding the
XX protein sequences in AB857020 to AB857374) or by determining the
XX expression profile of a gene group comprising these genes. The expression
XX levels or expression profiles produced by these genes are used as an
XX indicator when screening for ischaemic condition-improving drugs or
XX therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR

CC primers for a mouse ischaemic condition related sequence, which are used
CC in the exemplification of the present invention
XX
XX Sequence 10733 BP; 2166 A; 2890 C; 2968 G; 2704 T; 0 U; 5 Other;
SQ Query Match 68.8%; Score 17.2; DB 6; Length 10733;
Best Local Similarity 86.4%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 CGAGGGCCAGAGGCACATAG 24
DB 6259 CGAGGGCCAGAGGCACAGAG 6237
RESULT 23
AAK89507/C
ID AAK89507 standard; DNA; 16038 BP.
XX
XX AC AAK89507;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human digestive system antigen genomic sequence SEQ ID NO: 3083.
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
XX
XX WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001324.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180828P.
XX 24-FEB-2000; 2000US-0184564P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-021680P.
XX 11-JUL-2000; 2000US-0217487P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 18-AUG-2000; 2000US-0225759P.
XX 22-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 23-AUG-2000; 2000US-0227182P.
XX 30-AUG-2000; 2000US-0227009P.
XX 01-SEP-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.

```
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 20-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244741P.
PR 08-NOV-2000; 2000US-0244751P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WFI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
diagnosing, treating, preventing and/or prognosing disorders of the
digestive system, particularly cancer and cancer metastases.
PS Disclosure; SEQ ID NO 3083; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
number of human digestive system antigens. These can be used in the
diagnosis, treatment and prevention of digestive system disorders,
including cancer, Meckel's diverticulum, bacterial or parasitic
infections, appendicitis, Hirschsprung's disease, chronic colitis or
ulcerative colitis. The present sequence is a genomic DNA fragment
encoding a digestive system antigen of the invention
XX
SQ Sequence 16038 BP; 3380 A; 4425 C; 4399 G; 3834 T; 0 U; 0 Other;
Query Match 68.8%; Score 17.2; DB 4; Length 16038;
Best Local Similarity 86.4%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 4 GGAGGCCACAGGGGCACATAGC 25
Db 12626 GCAGGGTCAGAGGGCAGAGC 12605
RESULT 24
AAX23521
ID AAX23521 standard; DNA; 16595 BP.
XX
XX AC AAX23521;
XX
XX DT 17-JUN-1999 (first entry)
XX
XX Human kidney aminopeptidase P genomic DNA fragment 5.
XX
XX Aminopeptidase; human; Amp; gene therapy; treatment; Amp-deficiency;
prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;
arterial stenosis; industrial protein feed; malabsorption syndrome;
proteinaceous waste degradation; additive; immunohistochemistry; ss.
```

PR 26-DEC-2001; 2001US-00035832.
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
PI Morris DW;
XX
XX WPI; 2003-587068/55.
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1025; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed murine CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 23070 BP; 5504 A; 5904 C; 5830 G; 5559 T; 0 U; 273 Other;
Query Match 68.8%; Score 17.2; DB 8; Length 23070;
Best Local Similarity 86.4%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CGGAGGGCCAGAGGGCACATA 23
DB 18715 CTGGAGTGTAGAGGGCACATA 18736
RESULT 26
ADB72245
ID ADB72245 standard; DNA; 23070 BP.
XX
XX ADB72245;
AC
XX
DT 04-DEC-2003 (first entry)
XX
XX Mouse Wnt1 gene.
XX
XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX Mus sp.
OS
XX
XX WO2003008583-A2.
PN
XX
XX 30-JAN-2003.
PD
XX
XX 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
PR

OS Homo sapiens.
XX
XX WO9911799-A2.
PN
XX
XX 11-MAR-1999.
PD
XX
XX 02-SEP-1998; 98WO-US018426.
XX
XX 02-SEP-1997; 97US-0057854P.
XX
XX (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.
PA
XX
XX Ryan JW, Sprinkle TJC, Venema RC;
PI WPI; 1999-205193/17.
XX
XX Nucleic acid encoding human aminopeptidase P.
PT
XX
XX Claim 13; Page 192-201; 201pp; English.
XX
XX This invention describes the isolation of a novel human aminopeptidase P
CC (AmpP). This protein is used to produce recombinant Amp and can be used
CC for gene therapy for treating Amp-deficiency conditions. Its fragments
CC are used as primers and probes to identify patients with homozygous and
CC heterozygous Amp deficiency, including prenatal diagnosis (patients
CC defective in Amp are at risk of developing angioedema if treated with
CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors
CC in cases of excessive Amp expression. The product of the invention is
CC also used to identify Amp-expressing sequences in other animals and to
CC generate transgenic animals, and comparisons of genomic sequences are
CC used to detect mutations. Amp inhibitors are potentially useful as
CC antihypertensive agents and to prevent or treat arterial (re)stenosis or
CC atherosclerosis. The structure of Amp is used to design synthetic
CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal
CC imido bonds, can be used to degrade industrial protein feeds to free
CC amino acids, to degrade proteinaceous wastes, as additives in enzyme
CC formulations used to treat malabsorption syndrome and for studying its
CC biological role. Antibodies against Amp are used in immunohistochemical
CC methods to study Amp distribution
XX
SQ Sequence 16595 BP; 4429 A; 4145 C; 4168 G; 3853 T; 0 U; 0 Other;
Query Match 68.8%; Score 17.2; DB 2; Length 16595;
Best Local Similarity 86.4%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 GGAGGGCCAGAGGGCACATAGC 25
DB 12850 GGAGGGCCAGAGGGCCCAAGC 12871
RESULT 25
ADA02507
ID ADA02507 standard; DNA; 23070 BP.
XX
XX
XX ADA02507;
AC
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Mouse Wnt1 carcinoma associated gene, SEQ ID NO:1025.
DE
XX
XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW Gene; ds.
XX
XX Mus sp.
OS
XX
XX WO2003057146-A2.
PN
XX
XX 17-JUL-2003.
PD
XX
XX 26-DEC-2002; 2002WO-US041414.
PR
XX

XX (SAGR-) SAGRES DISCOVERY.
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-239337/23.
XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
PS Claim 1; SEQ ID NO 73; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ Sequence 23070 BP; 5504 A; 5904 C; 5830 G; 5559 T; 0 U; 273 Other;
Query Match 68.8%; Score 17.2; DB 9; Length 23070;
Best Local Similarity 86.4%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CGGAGGGCCAGAGGGCACATA 23
DB 18715 CTGGAGTGCTAGAGGGCACATA 18736
RESULT 27
ID AAF18390/c
XX AAF18390 standard; DNA; 376 BP.
XX
AC AAF18390;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 409.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW Cardioactive; immunomodulatory; muscular active; vulnerary;
KW Gastrointestinal; nephrotropic; antiinfective; gynecological;
KW Antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW Proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
FN WO20005180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US005918.
XX
PR 12-MAR-1999; 99US-0124270P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
DR P-PSDB; AAB5814.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.
XX
PS Claim 1; Page 877; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
CC used in the course of the invention for the identification and
CC characterisation of the polynucleotide and protein sequences
XX
SQ Sequence 376 BP; 78 A; 93 C; 118 G; 77 T; 0 U; 10 Other;
Query Match 68.0%; Score 17; DB 3; Length 376;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 GAGGGCCAGAGGGCACCA 21
DB 143 GAGGGCCAGAGGGCACCA 127
RESULT 28
ID ACH19776/c
XX ACH19776 standard; cDNA; 406 BP.
XX
AC ACH19776;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult lung cDNA #779.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
FN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 6988; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for

Qy 1 CCGGAGGSCACAGGGCACATAGC 25
||| ||| ||| ||| |||
Db 386 CCCGAAAGGAAGGACACATAGC 362

Search completed: June 20, 2004, 10:18:06
Job time : 160.776 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1168.25 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25
Sequence: 1 cccggaggccagggcacatagc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum Match 100%
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estcin.*
- 4: em_estm.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hic.*
- 9: gb_estl.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gssl.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| C 1 | 19.2 | 76.8 | 675 | 10 | BE309187 |
| C 2 | 19.2 | 76.8 | 861 | 14 | CF242277 |
| C 3 | 18.8 | 75.2 | 452 | 12 | BJ036626 |
| C 4 | 18.8 | 75.2 | 637 | 10 | AW645104 |

| | | | | | |
|------|------|------|------|----|----------|
| 5 | 18.8 | 75.2 | 760 | 28 | AZ189837 |
| 6 | 18.8 | 75.2 | 768 | 28 | BH312135 |
| 7 | 18.8 | 75.2 | 866 | 10 | BE248203 |
| 8 | 18.8 | 75.2 | 866 | 14 | CE587782 |
| C 9 | 18.6 | 74.4 | 236 | 14 | CD215458 |
| C 10 | 18.6 | 74.4 | 303 | 28 | AZ726396 |
| C 11 | 18.6 | 74.4 | 326 | 13 | BY124227 |
| C 12 | 18.6 | 74.4 | 351 | 13 | BY008855 |
| C 13 | 18.6 | 74.4 | 358 | 10 | BB800951 |
| C 14 | 18.6 | 74.4 | 584 | 12 | BM540159 |
| C 15 | 18.6 | 74.4 | 625 | 12 | BM538550 |
| C 16 | 18.6 | 74.4 | 637 | 10 | BM57042 |
| C 17 | 18.6 | 74.4 | 651 | 12 | BI652362 |
| C 18 | 18.6 | 74.4 | 734 | 10 | BF215493 |
| C 19 | 18.6 | 74.4 | 734 | 29 | CNS018K0 |
| C 20 | 18.6 | 74.4 | 825 | 12 | BI109078 |
| C 21 | 18.6 | 74.4 | 839 | 12 | BI646763 |
| C 22 | 18.6 | 74.4 | 945 | 29 | CNS020RM |
| C 23 | 18.6 | 74.4 | 1050 | 10 | BE887139 |
| C 24 | 18.6 | 74.4 | 3251 | 11 | AK051753 |
| C 25 | 18.4 | 73.6 | 350 | 28 | BZ344555 |
| C 26 | 18.4 | 73.6 | 587 | 29 | CE247615 |
| C 27 | 18.4 | 73.6 | 1201 | 9 | AL565528 |
| C 28 | 18.2 | 72.8 | 275 | 10 | BA483005 |
| C 29 | 18.2 | 72.8 | 446 | 13 | BY556125 |
| C 30 | 18.2 | 72.8 | 466 | 12 | BI243418 |
| C 31 | 18.2 | 72.8 | 478 | 9 | AI388401 |
| C 32 | 18.2 | 72.8 | 503 | 12 | BI167064 |
| C 33 | 18.2 | 72.8 | 527 | 28 | BZ860665 |
| C 34 | 18.2 | 72.8 | 541 | 28 | AZ478773 |
| C 35 | 18.2 | 72.8 | 576 | 9 | AI135996 |
| C 36 | 18.2 | 72.8 | 578 | 9 | AI108894 |
| C 37 | 18.2 | 72.8 | 578 | 9 | AI514125 |
| C 38 | 18.2 | 72.8 | 583 | 9 | AI106672 |
| C 39 | 18.2 | 72.8 | 586 | 9 | AI106827 |
| C 40 | 18.2 | 72.8 | 588 | 9 | AI063819 |
| C 41 | 18.2 | 72.8 | 594 | 9 | AI258372 |
| C 42 | 18.2 | 72.8 | 595 | 12 | BI589463 |
| C 43 | 18.2 | 72.8 | 601 | 28 | AQ763553 |
| C 44 | 18.2 | 72.8 | 602 | 12 | BG636173 |
| C 45 | 18.2 | 72.8 | 611 | 9 | AI064563 |
| C 46 | 18.2 | 72.8 | 619 | 12 | BI33912 |
| C 47 | 18.2 | 72.8 | 624 | 12 | BI376192 |
| C 48 | 18.2 | 72.8 | 629 | 12 | BI354430 |
| C 49 | 18.2 | 72.8 | 630 | 12 | BI168276 |
| C 50 | 18.2 | 72.8 | 640 | 12 | BI358748 |
| C 51 | 18.2 | 72.8 | 640 | 29 | CE734784 |
| C 52 | 18.2 | 72.8 | 643 | 9 | AI405779 |
| C 53 | 18.2 | 72.8 | 648 | 9 | AI135368 |
| C 54 | 18.2 | 72.8 | 652 | 12 | BI642035 |
| C 55 | 18.2 | 72.8 | 657 | 9 | AI064545 |
| C 56 | 18.2 | 72.8 | 657 | 12 | BI487260 |
| C 57 | 18.2 | 72.8 | 701 | 29 | CE417348 |
| C 58 | 18.2 | 72.8 | 733 | 29 | CNS01HHJ |
| C 59 | 18.2 | 72.8 | 748 | 9 | AA803450 |
| C 60 | 18.2 | 72.8 | 869 | 29 | CNS04UQ |
| C 61 | 18.2 | 72.8 | 976 | 10 | BF315659 |
| C 62 | 18.2 | 72.8 | 1007 | 29 | CNS011NR |
| C 63 | 18.2 | 72.8 | 343 | 13 | BI103079 |
| C 64 | 18.2 | 72.8 | 527 | 29 | BI122852 |
| C 65 | 18.2 | 72.8 | 672 | 29 | AL797485 |
| C 66 | 18.2 | 72.8 | 745 | 29 | BX203153 |
| C 67 | 18.2 | 72.8 | 800 | 29 | BI182875 |
| C 68 | 18.2 | 72.8 | 894 | 10 | BF234179 |
| C 69 | 18.2 | 72.8 | 1190 | 13 | BI381568 |
| C 70 | 17.8 | 71.2 | 237 | 10 | AW855329 |
| C 71 | 17.8 | 71.2 | 263 | 29 | CE197537 |
| C 72 | 17.8 | 71.2 | 395 | 12 | BI685011 |
| C 73 | 17.8 | 71.2 | 425 | 29 | CG384508 |
| C 74 | 17.8 | 71.2 | 486 | 29 | DR12U2S |
| C 75 | 17.8 | 71.2 | 491 | 12 | BI772144 |
| C 76 | 17.8 | 71.2 | 491 | 29 | CC704797 |
| C 77 | 17.8 | 71.2 | 495 | 29 | CC682864 |

| | | | |
|----------|-----------|------|----|
| AZ189837 | SP | 1015 | A |
| BH312135 | CH230 | -101 | |
| BF248203 | 601821 | 202 | |
| CE587782 | AGENCOURT | | |
| CD215458 | PPpin | PKO | |
| AZ726396 | RPCI | -24 | -1 |
| BY124227 | BY124227 | | |
| BY008855 | BY008855 | | |
| BB800951 | BB800951 | | |
| BM540159 | hb18g12 | g | |
| BM538550 | ha9ra10 | g | |
| BM57042 | BB657042 | | |
| BI652362 | 603299633 | | |
| BF215493 | 601881056 | | |
| AL164025 | Tetraodon | | |
| BI109078 | 602896924 | | |
| BI646763 | 603279793 | | |
| AL13187 | Tetraodon | | |
| BE887139 | 601508337 | | |
| AK051753 | Mus muscu | | |
| BZ344555 | hg87c01 | b | |
| CE247615 | tigr-gss- | | |
| AL565528 | AL565528 | | |
| BA483005 | BA483005 | | |
| BY556125 | BY556125 | | |
| BI243418 | RE41056 | 5 | |
| AI388401 | GH19304 | 5 | |
| BI167064 | RE07351 | 5 | |
| BZ860665 | CH24 | 256 | |
| AZ478773 | 1M029GK06 | | |
| AI135996 | GH13981 | 5 | |
| AI108894 | GH0206 | 5 | |
| AI514125 | GH27186 | 5 | |
| AI106672 | GH06745 | 5 | |
| AI106827 | GH06061 | 5 | |
| AI063819 | GH03910 | 5 | |
| AI258372 | LP01552 | 5 | |
| BI589463 | RH23274 | 5 | |
| AQ763553 | HS | 3156 | E |
| BG636173 | SD13773 | 5 | |
| AI064563 | GH05023 | 5 | |
| AI133912 | GH10955 | 5 | |
| BI376192 | RE62049 | 5 | |
| BI354430 | GM16474 | 5 | |
| BI168276 | RE08877 | 5 | |
| BI358748 | RE37095 | 5 | |
| CE734784 | tigr-gss- | | |
| AI405779 | GH25908 | 5 | |
| AI135368 | GH13067 | 5 | |
| BI642035 | SD24007 | 5 | |
| AI064545 | GH05002 | 5 | |
| BI487260 | RE0726 | 5 | |
| CE417348 | tigr-gss- | | |
| AL14440 | Anopheles | | |
| AA803450 | GM13207 | 5 | |
| AL293291 | Tetraodon | | |
| BF315659 | 601899312 | | |
| AL145976 | Anopheles | | |
| BI103079 | EX103079 | | |
| BI122852 | Danio rer | | |
| AL797485 | Danio rer | | |
| BX203153 | Danio rer | | |
| BI182875 | Danio rer | | |
| BF234179 | 602026039 | | |
| BI381568 | EX381568 | | |
| AW855329 | FW1-CT026 | | |
| CE197537 | tigr-gss- | | |
| BI685011 | 603307683 | | |
| CG384508 | OGVGA15TV | | |
| DR12U2S | Danio rer | | |
| BI772144 | 603054355 | | |
| CC704797 | OGVBM53TV | | |
| CC682864 | OGUL581V | | |

| | | | | | | | | | | | | | |
|-------|------|------|------|----|-----------|--------------------|-------|------|------|------|----|----------|----------------------------|
| C 78 | 17.8 | 71.2 | 506 | 10 | B8638537 | 946010D06 | 151 | 17.6 | 70.4 | 938 | 13 | BU432132 | 603257387 |
| C 79 | 17.8 | 71.2 | 549 | 28 | AQ529507 | RPCI-11-3 | 152 | 17.6 | 70.4 | 981 | 29 | CNS04NL8 | AL298853 Tetraodon |
| C 80 | 17.8 | 71.2 | 593 | 29 | CC692742 | OGUHE59TV | C 153 | 17.6 | 70.4 | 984 | 29 | CQ295025 | CG29AO05TV |
| C 81 | 17.8 | 71.2 | 613 | 29 | BX126344 | Danio rer | 154 | 17.6 | 70.4 | 1024 | 29 | CNS05CR5 | AL331466 Tetraodon |
| C 82 | 17.8 | 71.2 | 646 | 29 | BX226140 | Danio rer | C 155 | 17.6 | 70.4 | 1025 | 28 | B2179185 | BZ179185 CH230-349 |
| C 83 | 17.8 | 71.2 | 648 | 29 | BX125831 | Danio rer | C 156 | 17.6 | 70.4 | 1043 | 29 | CNS05S74 | AL351481 Tetraodon |
| C 84 | 17.8 | 71.2 | 655 | 29 | BX162646 | Danio rer | C 157 | 17.6 | 70.4 | 1047 | 29 | CNS0540B | AL320132 Tetraodon |
| C 85 | 17.8 | 71.2 | 661 | 29 | BX230269 | Danio rer | C 158 | 17.6 | 70.4 | 1121 | 14 | CS583237 | CS583237 EST002912 |
| C 86 | 17.8 | 71.2 | 709 | 28 | BX224716 | Danio rer | C 159 | 17.6 | 70.4 | 1126 | 12 | BM562134 | BM562134 AGENCOURT |
| C 87 | 17.8 | 71.2 | 722 | 28 | B2247146 | CH230-419 | 160 | 17.6 | 70.4 | 1726 | 11 | AK075847 | AK075847 Mus muscu |
| C 88 | 17.8 | 71.2 | 731 | 29 | BX147008 | Danio rer | 161 | 17.6 | 70.4 | 1731 | 10 | B2202272 | B2202272 BF17741 601882568 |
| C 89 | 17.8 | 71.2 | 737 | 14 | CA604722 | w1.pk004 | 162 | 17.4 | 69.6 | 237 | 10 | BF606195 | BF606195 273167 MA |
| C 90 | 17.8 | 71.2 | 745 | 29 | BX160297 | Danio rer | C 163 | 17.4 | 69.6 | 297 | 10 | BF606195 | BF606195 273167 MA |
| C 91 | 17.8 | 71.2 | 771 | 29 | BX189398 | Danio rer | C 164 | 17.4 | 69.6 | 441 | 14 | CS451304 | CS451304 C0603F09- |
| C 92 | 17.8 | 71.2 | 774 | 29 | BX161412 | Danio rer | C 165 | 17.4 | 69.6 | 463 | 28 | AQ885822 | AQ885822 HS 5527 B |
| C 93 | 17.8 | 71.2 | 845 | 28 | B2835651 | B2835651 CH240-209 | C 166 | 17.4 | 69.6 | 481 | 14 | CB728441 | CB728441 AMGNNUC:N |
| C 94 | 17.8 | 71.2 | 866 | 12 | BG445381 | GA_EA002 | C 167 | 17.4 | 69.6 | 583 | 28 | AZ995132 | AZ995132 2M0280A20 |
| C 95 | 17.8 | 71.2 | 1128 | 10 | BF672836 | 602152675 | C 168 | 17.4 | 69.6 | 603 | 14 | CA541739 | CA541739 C0609H03- |
| C 96 | 17.8 | 71.2 | 1201 | 13 | BX353858 | BX353858 | C 169 | 17.4 | 69.6 | 624 | 14 | CB578383 | CB578383 AMGNNUC:N |
| C 97 | 17.8 | 71.2 | 1201 | 13 | BX402018 | BX402018 | C 170 | 17.4 | 69.6 | 672 | 29 | AG135212 | AG135212 Pan trogl |
| C 98 | 17.6 | 70.4 | 185 | 28 | A2466655 | IM0277E03 | C 171 | 17.4 | 69.6 | 718 | 14 | CS536153 | CS536153 UI-M-G10- |
| C 99 | 17.6 | 70.4 | 260 | 10 | AW426057 | AW426057 59426 MAR | C 172 | 17.4 | 69.6 | 743 | 29 | BX136993 | BX136993 Danio rer |
| C 100 | 17.6 | 70.4 | 275 | 9 | AW783914 | AW783914 c8h11a1.f | C 173 | 17.4 | 69.6 | 759 | 12 | BM010948 | BM010948 603634592 |
| C 101 | 17.6 | 70.4 | 294 | 12 | BG401235 | BG401235 602465333 | C 174 | 17.4 | 69.6 | 808 | 29 | CC994191 | CC994191 ZUAAAX73TH |
| C 102 | 17.6 | 70.4 | 296 | 10 | B1943338 | B1943338 | C 175 | 17.4 | 69.6 | 838 | 29 | CG121728 | CG121728 PUFTH87TB |
| C 103 | 17.6 | 70.4 | 298 | 29 | CS368901 | CS368901 tigr-gss- | C 176 | 17.4 | 69.6 | 858 | 29 | CG088296 | CG088296 PUFTH87TB |
| C 104 | 17.6 | 70.4 | 343 | 9 | AA555977 | AA555977 v845b09.x | C 177 | 17.4 | 69.6 | 904 | 29 | CG121729 | CG121729 PUFTH87TD |
| C 105 | 17.6 | 70.4 | 359 | 10 | B8685008 | B8685008 | C 178 | 17.4 | 69.6 | 924 | 13 | BU147085 | BU147085 AGENCOURT |
| C 106 | 17.6 | 70.4 | 364 | 28 | AQ235341 | HS_2055_A | C 179 | 17.4 | 69.6 | 935 | 10 | BE970642 | BE970642 601679871 |
| C 107 | 17.6 | 70.4 | 403 | 10 | B8689841 | B8689841 | C 180 | 17.4 | 69.6 | 940 | 9 | AL532208 | AL532208 AGENCOURT |
| C 108 | 17.6 | 70.4 | 405 | 10 | B9463110 | UI-M-BZ0- | C 181 | 17.4 | 69.6 | 960 | 14 | CS570769 | CS570769 Drosophil |
| C 109 | 17.6 | 70.4 | 408 | 10 | B9693108 | B9693108 | C 182 | 17.4 | 69.6 | 1101 | 29 | CNS00HWN | AK073834 Mus muscu |
| C 110 | 17.6 | 70.4 | 412 | 9 | AI454036 | AI454036 v845b09.y | C 183 | 17.4 | 69.6 | 1101 | 11 | AK040162 | AK040162 Mus muscu |
| C 111 | 17.6 | 70.4 | 414 | 10 | B9592982 | B9592982 UI-M-BH4- | C 184 | 17.2 | 68.8 | 141 | 29 | CC586171 | CC586171 CH240-383 |
| C 112 | 17.6 | 70.4 | 418 | 13 | BX637670 | BX637670 | C 185 | 17.2 | 68.8 | 219 | 29 | CC586171 | CC586171 tigr-gss- |
| C 113 | 17.6 | 70.4 | 420 | 10 | AW120563 | AW120563 UI-M-BH1- | C 186 | 17.2 | 68.8 | 220 | 12 | BM755679 | BM755679 K-EST0033 |
| C 114 | 17.6 | 70.4 | 428 | 10 | AW124039 | AW124039 UI-M-BH2- | C 187 | 17.2 | 68.8 | 280 | 9 | AV330466 | AV330466 AV330466 |
| C 115 | 17.6 | 70.4 | 437 | 10 | B3684219 | B3684219 | C 188 | 17.2 | 68.8 | 280 | 9 | AV330466 | AV330466 AV330466 |
| C 116 | 17.6 | 70.4 | 440 | 9 | AW060424 | AW060424 UI-M-BH1- | C 189 | 17.2 | 68.8 | 280 | 9 | AV330466 | AV330466 AV330466 |
| C 117 | 17.6 | 70.4 | 452 | 10 | B2157245 | B2157245 RC4-HT037 | C 190 | 17.2 | 68.8 | 280 | 9 | AV330466 | AV330466 AV330466 |
| C 118 | 17.6 | 70.4 | 456 | 10 | AW093927 | AW093927 uf47d01.x | C 191 | 17.2 | 68.8 | 280 | 9 | AV330466 | AV330466 AV330466 |
| C 119 | 17.6 | 70.4 | 468 | 10 | BF011846 | BF011846 us42b01.y | C 192 | 17.2 | 68.8 | 322 | 13 | BY520186 | BY520186 BY520186 |
| C 120 | 17.6 | 70.4 | 472 | 10 | BF776862 | BF776862 288604 MA | C 193 | 17.2 | 68.8 | 332 | 10 | AK354460 | AK354460 36297 MAR |
| C 121 | 17.6 | 70.4 | 476 | 10 | BF769410 | BF769410 | C 194 | 17.2 | 68.8 | 335 | 29 | CC543231 | CC543231 CH240-424 |
| C 122 | 17.6 | 70.4 | 494 | 12 | BG689205 | BG689205 337113 BA | C 195 | 17.2 | 68.8 | 360 | 12 | BM755679 | BM755679 K-EST0033 |
| C 123 | 17.6 | 70.4 | 495 | 9 | AV111570 | AV111570 | C 196 | 17.2 | 68.8 | 381 | 9 | AV649893 | AV649893 AV649893 |
| C 124 | 17.6 | 70.4 | 499 | 12 | BM921509 | BM921509 AGENCOURT | C 197 | 17.2 | 68.8 | 381 | 9 | AA560373 | AA560373 v1390407.r |
| C 125 | 17.6 | 70.4 | 510 | 29 | PA687L | PA687L | C 198 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 126 | 17.6 | 70.4 | 528 | 14 | CB440511 | CB440511 690800 MA | C 199 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 127 | 17.6 | 70.4 | 529 | 9 | AV396810 | AV396810 | C 200 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 128 | 17.6 | 70.4 | 540 | 13 | BX525039 | BX525039 | C 201 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 129 | 17.6 | 70.4 | 544 | 28 | B2103023 | B2103023 CH230-147 | C 202 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 130 | 17.6 | 70.4 | 546 | 10 | BF041030 | BF041030 BP250018B | C 203 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 131 | 17.6 | 70.4 | 550 | 9 | AI791966 | AI791966 oq93807.y | C 204 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 132 | 17.6 | 70.4 | 552 | 14 | CB429212 | CB429212 605013 MA | C 205 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 133 | 17.6 | 70.4 | 626 | 10 | BE743077 | BE743077 601575132 | C 206 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 134 | 17.6 | 70.4 | 641 | 29 | CE201635 | CE201635 tigr-gss- | C 207 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 135 | 17.6 | 70.4 | 649 | 14 | CB443388 | CB443388 694266 MA | C 208 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 136 | 17.6 | 70.4 | 652 | 29 | CE134385 | CE134385 tigr-gss- | C 209 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 137 | 17.6 | 70.4 | 656 | 14 | CB245965 | CB245965 UI-M-F00- | C 210 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 138 | 17.6 | 70.4 | 677 | 14 | CA753584 | CA753584 002200190 | C 211 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 139 | 17.6 | 70.4 | 804 | 28 | CS375846 | CS375846 PUFTH87TB | C 212 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 140 | 17.6 | 70.4 | 811 | 28 | B2216631 | B2216631 | C 213 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 141 | 17.6 | 70.4 | 812 | 12 | BG611061 | BG611061 602612005 | C 214 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 142 | 17.6 | 70.4 | 834 | 29 | CNS025M5 | AL182246 Tetraodon | C 215 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 143 | 17.6 | 70.4 | 853 | 10 | BE036253 | BE036253 MO21E11 M | C 216 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 144 | 17.6 | 70.4 | 860 | 28 | B2547678 | B2547678 601076264 | C 217 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 145 | 17.6 | 70.4 | 860 | 28 | B2805951 | B2805951 PUFTH87TB | C 218 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 146 | 17.6 | 70.4 | 861 | 28 | B2805946 | B2805946 | C 219 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 147 | 17.6 | 70.4 | 897 | 14 | CB9840931 | CB9840931 | C 220 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 148 | 17.6 | 70.4 | 909 | 13 | B0642455 | B0642455 AGENCOURT | C 221 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 149 | 17.6 | 70.4 | 915 | 13 | B0439446 | B0439446 | C 222 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |
| C 150 | 17.6 | 70.4 | 935 | 10 | BF308856 | BF308856 601889783 | C 223 | 17.2 | 68.8 | 415 | 10 | BE723040 | BE723040 192375 MA |

| | | | | | | |
|------------|--|---------------|--------------|--------------------------|----------|------------------------|
| 297 | 17 | 68.0 | 234 | 28 | AZ397194 | 1M0162M01 |
| 298 | 17 | 68.0 | 270 | 29 | CB552473 | tigr-gss- |
| C 299 | 17 | 68.0 | 274 | 10 | BE070638 | CMA-BT041 |
| C 300 | 17 | 68.0 | 280 | 9 | AA317004 | AA317004 EST18869 |
| ALIGNMENTS | | | | | | |
| BE309187 | 601095072F1 | NCI_CGAP_Mam5 | Mus musculus | cdna clone IMAGE:3489664 | 5', | linear EST 26-OCT-2000 |
| BE309187 | mRNA sequence. | | | | | |
| BE309187 | mRNA sequence. | | | | | |
| BE309187 | EST. | | | | | |
| BE309187.1 | GI:9166997 | | | | | |
| BE309187 | Mus musculus (house mouse) | | | | | |
| BE309187 | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | | |
| BE309187 | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | | |
| BE309187 | 1 (bases 1 to 675) | | | | | |
| BE309187 | NIH-MGC http://img.nci.nih.gov/. | | | | | |
| BE309187 | National Institutes of Health, Mammalian Gene Collection (MGC) | | | | | |
| BE309187 | Unpublished (1999) | | | | | |
| BE309187 | Contact: Robert Strausberg, Ph.D. | | | | | |
| BE309187 | Email: cgaabs@mail.nih.gov | | | | | |
| BE309187 | Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys | | | | | |
| BE309187 | CDNA Library Preparation: Life Technologies, Inc. | | | | | |
| BE309187 | CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) | | | | | |
| BE309187 | DNA Sequencing by: Incyte Genomics, Inc. | | | | | |
| BE309187 | Clone distribution: MGC clone distribution information can be | | | | | |
| BE309187 | found through the I.M.A.G.E. Consortium/LLNL at: | | | | | |
| BE309187 | http://image.llnl.gov | | | | | |
| BE309187 | Plate: LLAM8530 row: p column: 17 | | | | | |
| BE309187 | High quality sequence stop: 522. | | | | | |
| BE309187 | Location/Qualifiers | | | | | |
| BE309187 | 1. .675 | | | | | |
| BE309187 | /organism="Mus musculus" | | | | | |
| BE309187 | /mol_type="mRNA" | | | | | |
| BE309187 | /strain="C57BL/6J" | | | | | |
| BE309187 | /db_xref="taxon:10090" | | | | | |
| BE309187 | /clone="IMAGE:3489664" | | | | | |
| BE309187 | /issue_type="tumor, gross tissue" | | | | | |
| BE309187 | /dev_stage="7 months" | | | | | |
| BE309187 | /lab_host="DH10B" | | | | | |
| BE309187 | /clone_lib="NCI CGAP Mam5" | | | | | |
| BE309187 | /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; | | | | | |
| BE309187 | Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. | | | | | |
| BE309187 | Library constructed by Life Technologies. Investigators | | | | | |
| BE309187 | providing samples: Lothar Hennighausen/Robin Humphreys, | | | | | |
| BE309187 | NIH" | | | | | |
| BE309187 | Query Match 76.8%; Score 19.2; DB 10; Length 675; | | | | | |
| BE309187 | Best Local Similarity 87.5%; Pred. No. 2.1e+03; | | | | | |
| BE309187 | Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0; | | | | | |
| BE309187 | Qy 1 CCGGAGGGCCAGAGGCACATAG 24 | | | | | |
| BE309187 | | | | | | |
| BE309187 | Db 516 CACGAGGGCCAGAGGCACCTAG 493 | | | | | |
| BE309187 | | | | | | |
| BE309187 | RESULT 2 | | | | | |
| BE309187 | CF242277/c | | | | | |
| BE309187 | LOCUS | | | | | |
| BE309187 | DEFINITION | | | | | |
| BE309187 | ACCESSION | | | | | |
| BE309187 | VERSION | | | | | |
| BE309187 | KEYWORDS | | | | | |
| BE309187 | SOURCE | | | | | |
| BE309187 | ORGANISM | | | | | |
| BE309187 | REFERENCE | | | | | |
| BE309187 | AUTHORS | | | | | |
| BE309187 | TITLE | | | | | |
| BE309187 | COMMENT | | | | | |
| BE309187 | AV610420 | | | | | |
| BE309187 | CE772891 tigr-gss- | | | | | |
| BE309187 | X1736229 Danio rer | | | | | |
| BE309187 | AL957092 AL957092 | | | | | |
| BE309187 | B1169645 R10883.5 | | | | | |
| BE309187 | BJ580769 BJ580769 | | | | | |
| BE309187 | AL862644 AL862644 | | | | | |
| BE309187 | Z86754 F.rubripes | | | | | |
| BE309187 | BE252309 BE252309 | | | | | |
| BE309187 | AL986525 Danio rer | | | | | |
| BE309187 | AQ310531 CITBI-E1- | | | | | |
| BE309187 | B1228645 R26254.5 | | | | | |
| BE309187 | BX229400 Danio rer | | | | | |
| BE309187 | B1485671 R568925.5 | | | | | |
| BE309187 | BX153639 Danio rer | | | | | |
| BE309187 | BX153639 Danio rer | | | | | |
| BE309187 | BH381917 AG-ND-173 | | | | | |
| BE309187 | EX144849 Danio rer | | | | | |
| BE309187 | BX207834 Danio rer | | | | | |
| BE309187 | BG340998 602462731 | | | | | |
| BE309187 | BX200167 Danio rer | | | | | |
| BE309187 | EX141348 Danio rer | | | | | |
| BE309187 | EX208901 Danio rer | | | | | |
| BE309187 | BX210325 Danio rer | | | | | |
| BE309187 | BX210325 Danio rer | | | | | |
| BE309187 | B2201364 CH230-445 | | | | | |
| BE309187 | BG175998 602335223 | | | | | |
| BE309187 | CE056270 tigr-gss- | | | | | |
| BE309187 | BF238803 601904402 | | | | | |
| BE309187 | BX396145 603805849 | | | | | |
| BE309187 | BX152564 Danio rer | | | | | |
| BE309187 | EX178813 Danio rer | | | | | |
| BE309187 | EX181121 Danio rer | | | | | |
| BE309187 | BF303575 602031192 | | | | | |
| BE309187 | CF281474 602401993 | | | | | |
| BE309187 | CF147554 AGENCOURT | | | | | |
| BE309187 | BM415551 OP20827 M | | | | | |
| BE309187 | CA968911 CcL03a10 | | | | | |
| BE309187 | B1736630 603360209 | | | | | |
| BE309187 | BU594379 AGENCOURT | | | | | |
| BE309187 | BG530583 602560171 | | | | | |
| BE309187 | AL199293 Tetradon | | | | | |
| BE309187 | BF679992 602154705 | | | | | |
| BE309187 | CF878053 trico16xe | | | | | |
| BE309187 | BU946652 703769886 | | | | | |
| BE309187 | BE858867 601507518 | | | | | |
| BE309187 | AL167880 Tetradon | | | | | |
| BE309187 | CA968251 CcLM07a2 | | | | | |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.

REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (Bases 1 to 861)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1A14629 row: e column: 09
High quality sequence stop: 668.

FEATURES source

Location/Qualifiers
1..861

/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:6976306"
/tissue_type="tailbud"
/dev_stage="embryo, stages 20-27"
/lab_hosts="DH10B (phage-resistant)"
/clone_lib="NICHD XGC Emb7"
/note="Vector: pCMV-SPORT6.1; Site1: NotI; Site2: EcoRV;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Invitrogen. Note: This is a
Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 861;
Best Local Similarity 87.5%; Pred. No. 2.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCCGAGGCGCCAGAGGCGACATAG 24
DB 861 CACGAGGCGCCAGTGGCGACAAAG 838

RESULT 3 BJ036626

LOCUS BJ036626 452 bp mRNA linear EST 26-SEP-2003
DEFINITION laevis cDNA clone XL036p06 5', mRNA sequence.
ACCESSION BJ036626
VERSION BJ036626.1 GI:17397211
KEYWORDS EST.

ORGANISM

Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.

REFERENCE AUTHORS

1 (Bases 1 to 452)
Kitayama A., Takasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
Kohara, Y.

TITLE JOURNAL COMMENT

Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehin@genes.nig.ac.jp
The information of this clone is available through the following
URL.

FEATURES source

<http://xenopus.nibb.ac.jp>.
Location/Qualifiers
1..452
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL036p06"
/tissue_type="whole embryo"
/dev_stage="stage 15"
/clone_lib="NIBB Mochii normalized Xenopus neurula
library"

ORIGIN

Query Match 75.2%; Score 18.8; DB 12; Length 452;
Best Local Similarity 90.9%; Pred. No. 2.7e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCGAGGCGCCAGGCGCACAT 22
DB 345 CTCGAGAGCCAGAGGCGCACAT 366

RESULT 4 AW645104

LOCUS AW645104 637 bp mRNA linear EST 26-APR-2001
DEFINITION cm48h12.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0147H12 5', mRNA sequence.

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AW645104
AW645104.1 GI:7402520
EST.
Xenopus laevis (African clawed frog)
Xenopus laevis

REFERENCE AUTHORS

1 (bases 1 to 637)
Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.
Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M.,
Touchman, J.W., Bonaldo, M.F. and Soares, M.B.
The NIEHS Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs

JOURNAL MEDLINE PUBMED COMMENT

Gene 267 (1), 71-87 (2001)
21211403
11311557
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA

Tel: 919 541-4899
Fax: 919 541-4571

Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com

DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).

PCR Primers
FORWARD: TGTAAACACGAGGCGCAGT
BACKWARD: CAGGAACAGCTATGACC
Plate: 0147 row: H column: 12
Seq primer: T7 primer.

FEATURES source

Location/Qualifiers
1..637
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="PBX0147H12"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"

/lab_host="DH10B"
/clone_lib="Blackshear/Soares normalized Xenopus egg library"
/note="vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI; PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dn18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10⁵ recombinants, with average insert sizes of 1-1.5 kb."

ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 637;
Best Local Similarity 90.9%; Pred. No. 3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCGGAGGCGCAGAGGCACAT 22
Db 360 CTCGAGAGCCAGAGGCACAT 381

RESULT 5

AZ189837
LOCUS
DEFINITION
SP_1015 A2_F12_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plates=1015 Col=24 Row=K, genomic survey sequence.

ACCESSION

AZ189837
VERSION
KEYWORDS
SOURCE
ORGANISM

Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE

Ameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)

TITLE

Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA

JOURNAL

Medline
PubMed
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1015 row: K column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 760.

FEATURES

source
location/Qualifiers
1..760
/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="Plates=1015 Col=24 Row=K"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 75.2%; Score 18.8; DB 28; Length 760;
Best Local Similarity 90.9%; Pred. No. 3.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 CCGAGGCGCAGAGGCACATAG 24
Db 600 CCGAGGCGCAGAGGCACATAG 621

ORIGIN

Query Match 75.2%; Score 18.8; DB 28; Length 760;
Best Local Similarity 90.9%; Pred. No. 3.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 CCGAGGCGCAGAGGCACATAG 24
Db 600 CCGAGGCGCAGAGGCACATAG 621

RESULT 6

BH312135
LOCUS
DEFINITION
CH230-101J1.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone CH230-101J1, genomic survey sequence.

ACCESSION

BH312135
VERSION
KEYWORDS
SOURCE
ORGANISM

Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 768)
Zhao, S., Shetty, J., Shatsman, S., Tsengaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment Unpublished (1999)

TITLE

Other_GSSs: CH230-101J1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

COMMENT

Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 101 row: J column: 1
Seq primer: T7
Class: BAC ends.

FEATURES

source
location/Qualifiers
1..768
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-101J1"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: PTABAC2.1; Site 1: EcoRI; Site 2: EcoRI; CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by Pieter de Jong"

ORIGIN

Query Match 75.2%; Score 18.8; DB 28; Length 768;
Best Local Similarity 90.9%; Pred. No. 3.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCGGAGGCGCAGAGGCACAT 22
Db 745 CCGGAGGCGCAGAGGCACAT 766

RESULT 7

BF248203
LOCUS
DEFINITION
601821202F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4046136 5',

```

mRNA sequence.
ACCESSION BF248203
VERSION BF248203.1 GI:11164466
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 866)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM873 row: c column: 01
High quality sequence stop: 402.
FEATURES
source
1..866
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4046136"
/tissue_type="melanotic melanoma, high MDR"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_62"
/Note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA.
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGCGCGGCGGCACATG-DT(30)BN-3',
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
ORIGIN
Query Match 75.2%; Score 18.8; DB 10; Length 866;
Best Local Similarity 90.9%; Pred.No. 3.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 CGAGGGCCAGGCGGCACATAG 24
Db 523 CGAGGGCCAGGCGGCACACAG 544
RESULT 8
CB587782
LOCUS CB587782
DEFINITION AGENCOURT 12974016 NIH_MGC 136 Mus musculus cDNA clone
IMAGE:30293489 5', mRNA sequence.
ACCESSION CB587782
VERSION CB587782.1 GI:29505638
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS 1 (bases 1 to 866)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe

cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM339 row: a column: 18
High quality sequence start: 16
High quality sequence stop: 485.
FEATURES
source
1..866
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30293489"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/Note="Vector: PCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCGGCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 75.2%; Score 18.8; DB 14; Length 866;
Best Local Similarity 90.9%; Pred.No. 3.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGGCGGCACAT 22
Db 664 CCCTGAGGGCCAGGCGGCACAT 685
RESULT 9
CD215458/c
LOCUS CD215458
DEFINITION pp1n.pk002.g9 Normalized chicken pituitary/hypothalamus/pineal
cDNA library (pp1n) Gallus gallus cDNA clone pp1n.pk002.g9 5',
similar to dbj|BAA92562.1 (AB037745) KIAA1324 protein [Homo
sapiens], mRNA sequence.
ACCESSION CD215458
VERSION CD215458.1 GI:30955457
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
AUTHORS 1 (bases 1 to 236)
TITLE Porter, T.E. and Cogburn, L.A.
JOURNAL Chicken ESTs from pituitary/hypothalamus/pineal gland
COMMENT Unpublished (2001)
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES
source
1..236
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="pp1n.pk002.g9"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"

```


Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source
Location/Qualifiers
1..326
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L630023J01"
/sex="male"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male brain"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCCGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using triazole thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTAAATTAATCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 326;
Best Local Similarity 84.0%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGGCACATAGC 25
DB 57 CCGGAGGGCCGCTAGGCACAGAGC 33

RESULT 12

LOCUS BY008855/c
DEFINITION BY008855 RIKEN full-length enriched, 16 days neonate male dienecephalon Mus musculus cDNA clone G63003K24 5', mRNA sequence.
ACCESSION BY008855
VERSION BY008855.1 GI:26069104
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 351)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Buit, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusica, V., Chotha, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,

Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Varado, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Akazawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
MEDLINE
PUBMED
22354683
12466851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome.res@gsc.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source
Location/Qualifiers
1..351
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G63003K24"
/sex="male"
/tissue_type="dienecephalon"
/dev_stage="16 days neonate"
/clone_lib="RIKEN full-length enriched, 16 days neonate male dienecephalon"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 351;
Best Local Similarity 84.0%; Pred. No. 3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGGCACATAGC 25

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE AUTHORS

O'Shaughnessy,A.L., McCombie,W.R., Baker,J.P., Balija,V.,
Cunning,D., Dedhia,N.N., de la Bastide,M., Katzenberger,F.,
King,L., Kirchoff,K.A., Miller,B., Muller,S., Nascimento,L.U.,
Palmer,L., Santos,L., Shah,R.S., Spiegel,L.A., Zuravarn,T.,
Preston,R. and Hamon,G.J.

TITLE JOURNAL

Expressed sequence tags from Canis familiaris (dog) (2002)

COMMENT

Unpublished (2002)
Contact: W. Richard McCombie
Lita Anenberg Hazen Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100 Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874

Email: mcombie@cshl.org
Plate: ha97 row: a column: 10
Seq primer: -21M3UnivRev
High quality sequence stop: 625.

FEATURES

source

```
1..625
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="ha97a10"
/tissue_type="testes"
/clone_lib="Canis cDNAs from testes cells"
/note="Vector: Lambda Zap II; The library was produced by Greg Hannon and Raymond Preston (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stratagene zap cDNA synthesis kit. It was made from dog testes. Please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."
```

ORIGIN

```
Query Match      74.4%; Score 18.6; DB 12; Length 625;
Best Local Similarity 84.0%; Pred. No. 3.6e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGCGCCAGAGGCACATAGC 25
Db 261 CCGGATGACCGAGTGGCAGATAGC 285
```

RESULT 16

BB657042/c

LOCUS

DEFINITION

BB657042 RIKEN full-length enriched, 12 days embryo spinal ganglion

Mus musculus cDNA clone D130072G11 5', mRNA sequence.

BB657042

VERSION

BB657042.1 GI:16490868

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 637)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,

Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,

Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,

Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,

Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Takeda,Y., Tanaka,T., Taya,T., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsukuba-Ku, Ibaraki, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,

Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.

and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,

Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,

Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

e mouse tissues.

Location/Qualifiers

1..637

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="D130072G11"

/tissue_type="spinal ganglion"

/dev_stage="12 days embryo"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 12 days embryo

spinal ganglion"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGCGCGCGCAACTCGAGTCTTTTCTTTT 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGAGATCTCGAGTCTTAATTAATTCCTCCCCCCCC 3']. cDNA

was cleaved with BamHI and XhoI. Vector: a modified

pluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

```
Query Match      74.4%; Score 18.6; DB 10; Length 637;
Best Local Similarity 84.0%; Pred. No. 3.6e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGCGCCAGAGGCACATAGC 25
Db 31 CCGGAGGCGCCGTAGGCGCACAGAGC 7
```

RESULT 17

BI652362/c

LOCUS

DEFINITION

BI652362 F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5340176 5',

mRNA sequence.

BI652362

VERSION

BI652362.1 GI:15566598

EST.

genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES

source
Location/Qualifiers
1..734

/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="168E09"
/clone_lib="G"
/note="Genoscope sequence ID : COAG168AC05LP1-end : T7"

ORIGIN

Query Match 74.4%; Score 18.6; DB 29; Length 734;
Best Local Similarity 84.0%; Pred. No. 3.8e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCGACAGGGCACATAGC 25
DB 454 CCCGTCGGGACAGAGCACATAGC 478

RESULT 20

LOCUS BI109078/c 825 bp mRNA linear EST 26-JUN-2001
DEFINITION 602896924F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5027164 5',
mRNA sequence.

ACCESSION BI109078.1 GI:14559979

VERSION

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 825)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Prepared by: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1076 row: 0 column: 05

High quality sequence start: 11

High quality sequence stop: 767.

FEATURES

source

1..825
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5027164"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN

Query Match 74.4%; Score 18.6; DB 12; Length 825;
Best Local Similarity 84.0%; Pred. No. 4e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCGACAGGGCACATAGC 25

DB 36 CCGGAGGCCGCTAGGGCACAGAGC 12

RESULT 21

LOCUS

DEFINITION

BI646763 839 bp mRNA linear EST 12-SEP-2001

603279793F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5320031 5',
mRNA sequence.

ACCESSION BI646763.1 GI:15560999

VERSION

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 839)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Prepared by: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1811 row: 1 column: 24

High quality sequence stop: 823.

FEATURES

source

1..839
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129, C57BL/6J, FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5320031"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam3"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."

ORIGIN

Query Match 74.4%; Score 18.6; DB 12; Length 839;

Best Local Similarity 84.0%; Pred. No. 4e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCGACAGGGCACATAGC 25

DB 820 CCCAAAGTCCAGAGGGCACAGAGC 796

RESULT 22

LOCUS

DEFINITION

CNS02THM 945 bp DNA linear GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
164107 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL213187

VERSION

KEYWORDS

SOURCE GSS; genome survey sequence.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.

| REFERENCE | AUTHORS | TITLE | JOURNAL | MEDLINE | PUBMED | FEATURES | source |
|-----------|--|--|------------------------------------|----------|---------------|-----------|---------|
| 1 | Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fzames,C., Wincker,P., Brottier,P., Quetier,F., Sautin,W. and Weissenbach,J. | Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence | Nat Genet. 25 (2), 235-238 (2000) | 20296633 | 10835645 | | |
| 2 | Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and Weissenbach,J. | Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis | Genome Res. 10 (7), 939-949 (2000) | 20359837 | 10899143 | | |
| 3 | (bases 1 to 945) | Direct Submission | Genoscope. | | | | |
| 4 | Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segrei@genoscope.cns.fr - Web : www.genoscope.cns.fr) | This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon. | | | | | |
| 5 | Location/Qualifiers | | | | | | |
| 6 | 1..945 | | | | | | |
| 7 | /organism="Tetraodon nigroviridis" | | | | | | |
| 8 | /mol_type="genomic DNA" | | | | | | |
| 9 | /db_xref="taxon:99883" | | | | | | |
| 10 | /clone="164107" | | | | | | |
| 11 | /clone_lib="G" | | | | | | |
| 12 | /note="Genoscope sequence ID : COAG164AE041P1-end : T7" | | | | | | |
| 13 | ORIGIN | | | | | | |
| 14 | Query Match | 74.4%; | Score 18.6; | DB 29; | Length 945; | | |
| 15 | Best Local Similarity | 84.0%; | Pred. No. 4.1e+03; | | | | |
| 16 | Matches | 21; | Conservative | 0; | Mismatches 4; | Indels 0; | Gaps 0; |
| 17 | QY | 1 | CCCGAGGGCCAGAGGCACATAGC | 25 | | | |
| 18 | DB | 692 | CCCGTCGGCCAGAGGCACATAGC | 716 | | | |
| 19 | RESULT 23 | | | | | | |
| 20 | BE887139 | | | | | | |
| 21 | LOCUS | BE887139.1 | GI:10342129 | | | | |
| 22 | DEFINITION | Homo sapiens (human) | | | | | |
| 23 | ACCESSION | BE887139 | | | | | |
| 24 | VERSION | BE887139.1 | GI:10342129 | | | | |
| 25 | KEYWORDS | EST. | | | | | |
| 26 | SOURCE | Homo sapiens | | | | | |
| 27 | ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| 28 | REFERENCE | NIH-MGC http://mgi.nci.nih.gov/. | | | | | |
| 29 | AUTHORS | 1 (bases 1 to 1050) | | | | | |
| 30 | TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) | | | | | |
| 31 | JOURNAL | Unpublished (1999) | | | | | |
| 32 | COMMENT | Contact: Robert Strausberg, Ph.D. Email: c9apbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov | | | | | |
| 33 | 1 column: 10 | | | | | | |
| 34 | High quality sequence stop: 733. | | | | | | |
| 35 | Location/Qualifiers | | | | | | |
| 36 | 1..1050 | | | | | | |
| 37 | /organism="Homo sapiens" | | | | | | |
| 38 | /mol_type="mRNA" | | | | | | |
| 39 | /db_xref="taxon:9606" | | | | | | |
| 40 | /clone="IMAGE:3909969" | | | | | | |
| 41 | /tissue_type="leiomyosarcoma" | | | | | | |
| 42 | /lab_host="DH10B (phage-resistant)" | | | | | | |
| 43 | /clone_lib="NIH MGC 71" | | | | | | |
| 44 | /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr. | | | | | | |
| 45 | Average insert size 2.1 kb. | | | | | | |
| 46 | ORIGIN | | | | | | |
| 47 | Query Match | 74.4%; | Score 18.6; | DB 10; | Length 1050; | | |
| 48 | Best Local Similarity | 84.0%; | Pred. No. 4.3e+03; | | | | |
| 49 | Matches | 21; | Conservative | 0; | Mismatches 4; | Indels 0; | Gaps 0; |
| 50 | QY | 1 | CCCGAGGGCCAGAGGCACATAGC | 25 | | | |
| 51 | DB | 721 | CCCGAGGGCCAGAGGCATGCG | 745 | | | |
| 52 | RESULT 24 | | | | | | |
| 53 | AK051753/c | | | | | | |
| 54 | LOCUS | AK051753.1 | GI:26342185 | | | | |
| 55 | DEFINITION | Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length enriched library, clone:DJ30072G11 product:hstone deacetylase 2, full insert sequence. | | | | | |
| 56 | ACCESSION | AK051753 | | | | | |
| 57 | VERSION | AK051753.1 | GI:26342185 | | | | |
| 58 | KEYWORDS | HTC; CAP trapper. | | | | | |
| 59 | SOURCE | Mus musculus (house mouse) | | | | | |
| 60 | ORGANISM | Mus musculus | | | | | |
| 61 | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | | | |
| 62 | REFERENCE | Carninci,P. and Hayashizaki,Y. | | | | | |
| 63 | AUTHORS | High-efficiency full-length cDNA cloning | | | | | |
| 64 | TITLE | Meth. Enzymol. 303, 19-44 (1999) | | | | </ | |

| | | | | | | | | | | |
|---------------------------------|---|--|--|--|--|--|--|--|--|--|
| JOURNAL REFERENCE AUTHORS | of 60,770 full-length cDNAs | | | | | | | | | |
| | Nature 420, 563-573 (2002) | | | | | | | | | |
| | 6 (bases 1 to 3251) | | | | | | | | | |
| | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y. | | | | | | | | | |
| TITLE JOURNAL | Direct Submission | | | | | | | | | |
| | Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) | | | | | | | | | |
| | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. | | | | | | | | | |
| | Please visit our web site for further details. URL: http://fantom.gsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/ | | | | | | | | | |
| FEATURES source | Location/Qualifiers | | | | | | | | | |
| | 1..3251 | | | | | | | | | |
| | /organism="Mus musculus" | | | | | | | | | |
| | /mol_type="mRNA" | | | | | | | | | |
| CDS | /strain="C57BL/6J" | | | | | | | | | |
| | /db_xref="FANTOM DB:D130072G11" | | | | | | | | | |
| | /db_xref="MGI:2420128" | | | | | | | | | |
| | /db_xref="taxon:10090" | | | | | | | | | |
| ORIGIN | /clone="D130072G11" | | | | | | | | | |
| | /tissue_type="spinal ganglion" | | | | | | | | | |
| | /clone_lib="RIKEN full-length enriched mouse cDNA library" | | | | | | | | | |
| | /dev_stage="12 days embryo" | | | | | | | | | |
| Query Match | 184..1098 | | | | | | | | | |
| | /note="unnamed protein product; histone deacetylase 2 (MGDI MG1:1097691, GB NM_008229, evidence: BLASTN, 99%, match=1011) | | | | | | | | | |
| | putative" | | | | | | | | | |
| | /codon_start=1 | | | | | | | | | |
| Best Local Similarity | /protein_id="BAC34755.1" | | | | | | | | | |
| | /db_xref="GI:26342186" | | | | | | | | | |
| | /translation="MAYSGGGKKVCYVDGIDGNYVGGQHPKPHRMTHNLLL NYGLYRKEIIVPHKTAETEMTKYSDIIRPLRSIRPDNMSEYKQMRPNVEDCP VDFGFEFCOLSTGDSVAGAVKLNRQDTMAVNWAGLHAKSEAGSCFVYNDIVLA ILELLKYHFORLVLYIDITHGIDGVEEAPYTDRTVMTVSFKYGFYFPGTGLRIDGAG KKGYAVNPNRMDGIDDSYGOIFKPIISKWENYQPSAVVLQGGADSLSDGLGCFN LTVKGRQTFPLVWCFSPTSHKSPHC" | | | | | | | | | |
| | 1 CCCGAGGGCCACAGGGCCATAGC 25 | | | | | | | | | |
| QY | | | | | | | | | | |
| | DB | | | | | | | | | |
| | 35 CCGGAGGGCCGTAGGGCCACAGC 11 | | | | | | | | | |
| | 74.4%; Score 18.6; DB 11; Length 3251; Query Match 84.0%; Pred. No. 6.3e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0; | | | | | | | | | |
| RESULT 25 | BZ344555 | | | | | | | | | |
| | BZ344555 | | | | | | | | | |
| | LOCUS | | | | | | | | | |
| | hg87c01.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum | | | | | | | | | |
| DEFINITION | bicolor genomic clone hg87c01 5', genomic survey sequence. | | | | | | | | | |
| | Accession | | | | | | | | | |
| | Version | | | | | | | | | |
| | Keywords | | | | | | | | | |
| SOURCE | Canis familiaris (dog) | | | | | | | | | |
| | Canis familiaris | | | | | | | | | |
| | Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. | | | | | | | | | |
| | 1 (bases 1 to 587) | | | | | | | | | |
| REFERENCE | Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C. | | | | | | | | | |
| | The dog genome: survey sequencing and comparative analysis | | | | | | | | | |
| | Science 301 (5641), 1898-1903 (2003) | | | | | | | | | |
| | JOURNAL | | | | | | | | | |
| TITLE | MEDLINE | | | | | | | | | |
| | PUBMED | | | | | | | | | |
| | 14512627 | | | | | | | | | |
| | Contact: Kirkness EF | | | | | | | | | |

The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

FEATURES
Location/Qualifiers
1..587

/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 73.6%; Score 18.4; DB 29; Length 587;
Best Local Similarity 95.0%; Pred. No. 4.3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGGCACA 21
DB 254 CCTGAGGCCAGAGGGGCACA 273

RESULT 27
LOCUS
DEFINITION AL565528 Homo sapiens FETAL BRAIN Homo sapiens cdna clone
CSODF004XJ16 3-PRIME, mRNA sequence.

ACCESSION
VERSION AL565528.2 GI:30549632

KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)

REFERENCE
1 (bases 1 to 1201)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE
Full-length cdna libraries and normalization
JOURNAL
Unpublished (2001)

COMMENT
On Feb 16, 2001 this sequence version replaced gi:12916994.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2308.f For
more information about this cluster, see
http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODF004DE08NP1&cluster=2308.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODF004DE08NP1.

FEATURES

Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF004XJ16"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"

/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cdna
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cdna was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 73.6%; Score 18.4; DB 9; Length 1201;

Best Local Similarity 58.3%; Pred. No. 5.4e+03;
Matches 14; Conservative 9; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCGAGGGCCAGAGGGGCACATAG 24
DB 62 SCWVRASSCCSRGAGGGGACATCG 39

RESULT 28

LOCUS
DEFINITION BB483005/c

musculus cdna clone D430018A20 3' similar to X06340 Mouse mRNA for
p-cadherin, mRNA sequence.

ACCESSION
VERSION BB483005
KEYWORDS
SOURCE EST.

ORGANISM
Mus musculus (house mouse)

REFERENCE
1 (bases 1 to 275)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,
Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,
Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,
Okazaki, Y., Otsu, T., Owa, C., Saito, H., Sakai, C., Sato, K.,
Shibata, Y., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,
Takahashi, F., Tomimaga, N., Tova, T., Tsunoda, Y., Watabiki, A.,
Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A.,
Yokota, T., Yoshida, K., Yoshino, A., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

TITLE
RIKEN Mouse ESTs (Konno, H., et al.)
JOURNAL
Unpublished (2000)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cdna. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Tomaru, Y., Carninci, P., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cdna cloning. Methods Enzymol. 303,
19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.
Location/Qualifiers
1..275
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="D430018A20"
/tissue_type="lung"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 13 days embryo
lung"

FEATURES

Location/Qualifiers
1..275
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="D430018A20"
/tissue_type="lung"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 13 days embryo
lung"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5].
 GAGAGAGAGCGCCGACACTCGAGTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5].
 GAGAGAGATTCGAGTAAATTAATATCCCTCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 275;
 Best Local Similarity 87.0%; Pred. No. 4e+03; 3; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 3

QY 2 CCGGAGGCGCCAGAGGCGACATAG 24

Db 53 CTGGAGGCGCCAGAGGCGAGATAG 31

RESULT 29

BY556125/c

LOCUS

DEFINITION

cells Mus musculus cDNA clone F730316D02 3', mRNA sequence.

ACCESSION

BY556125

VERSION

BY556125.1 GI:26890504

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY556125 446 bp mRNA linear EST 14-DEC-2002
 BY556125 RIKEN full-length enriched, B6-derived CD11 +ve dendritic
 cells Mus musculus cDNA clone F730316D02 3', mRNA sequence.
 BY556125
 BY556125.1 GI:26890504
 Mus musculus (house mouse)
 Mus musculus
 Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 446)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamakawa, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusci, V.,
 Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
 Sanderlin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp;
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC Building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES

source

1. 446
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="F730316D02"
 /cell_type="B6-derived CD11 +ve dendritic cells"
 /clone_lib="RIKEN full-length enriched, B6-derived CD11
 +ve dendritic cells"

ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 446;
 Best Local Similarity 87.0%; Pred. No. 4.7e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGCGCCAGAGGCGACATA 23

Db 249 CCGGAGGCGCCAGAGGCGACATA 227

RESULT 30

B1243418/c

LOCUS

DEFINITION

Drosophila melanogaster

mRNA sequence.

B1243418

B1243418.1

GI:114712301

EST.

KEYWORDS

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 466)

REFERENCE

AUTHORS

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Misra, S., Mungall, C. J., Nunco, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and
Rubin, G. M.
BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE03582: arm:2L [2452680,2747991]
estimated-cyto:22F3-23B3: 05/12/2001
Plate: RE.410 row: E column: 8
High quality sequence stop: 409.

FEATURES
source
1..466
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE41056"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DHS-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 72.8%; Score 18.2; DB 12; Length 466;
Best Local Similarity 87.0%; Pred. No. 4.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 CGGAGGGCCAGAGGGCACATAGC 25
DB 414 CGGATGGCCAGAGCACATTGC 392

Search completed: June 20, 2004, 14:15:35
Job time : 1194.25 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 27.6455 Seconds
(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccgagggccagagccatagc 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 300 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|--------------------|--------------------|
| 1 | 25 | 100.0 | 4403765 | US-09-103-840A-2 | Sequence 2, Appli |
| 2 | 25 | 100.0 | 4411529 | US-09-103-840A-1 | Sequence 1, Appli |
| 3 | 17.2 | 68.8 | 16595 | US-09-146-053-7 | Sequence 7, Appli |
| 4 | 16.8 | 67.2 | 481 | US-09-621-976-167 | Sequence 167, App |
| 5 | 16.8 | 67.2 | 2059 | US-09-604-978-5 | Sequence 5, Appli |
| 6 | 16.8 | 67.2 | 2059 | US-09-604-978-5 | Sequence 5, Appli |
| 7 | 16.6 | 66.4 | 533 | US-08-049-283A-1 | Sequence 1, Appli |
| 8 | 16.6 | 66.4 | 10684 | US-08-618-100B-3 | Sequence 3, Appli |
| 9 | 16.4 | 65.6 | 1494 | US-09-277-262-3 | Sequence 7, Appli |
| 10 | 16.4 | 65.6 | 2240 | US-09-277-262-7 | Sequence 7, Appli |
| 11 | 16.2 | 64.8 | 1262 | US-09-482-273-44 | Sequence 44, Appli |
| 12 | 16.2 | 64.8 | 1357 | US-09-448-176-9 | Sequence 9, Appli |
| 13 | 16.2 | 64.8 | 1357 | US-09-448-176-10 | Sequence 10, Appl |
| 14 | 16.2 | 64.8 | 1357 | US-09-448-176-11 | Sequence 11, Appl |
| 15 | 16.2 | 64.8 | 3867 | US-08-762-428A-5 | Sequence 5, Appli |
| 16 | 16.2 | 64.8 | 4016 | US-08-762-428A-7 | Sequence 7, Appli |
| 17 | 16 | 64.0 | 1278 | US-09-050-159-123 | Sequence 123, App |
| 18 | 16 | 64.0 | 1973 | US-09-976-994-193 | Sequence 193, App |
| 19 | 16 | 64.0 | 2201 | US-09-566-921-52 | Sequence 52, Appl |
| 20 | 16 | 64.0 | 3600 | US-09-462-917A-106 | Sequence 106, App |
| 21 | 16 | 64.0 | 17606 | US-08-943-731-4 | Sequence 4, Appli |
| 22 | 15.8 | 63.2 | 1231 | US-09-015-188-5 | Sequence 5, Appli |
| 23 | 15.8 | 63.2 | 7680 | US-09-210-748A-3 | Sequence 3, Appli |
| 24 | 15.8 | 63.2 | 49377 | US-08-764-233A-1 | Sequence 1, Appli |
| 25 | 15.6 | 62.4 | 539 | US-09-621-976-1421 | Sequence 1421, Ap |
| 26 | 15.6 | 62.4 | 1518 | US-09-719-919A-2 | Sequence 2, Appli |
| 27 | 15.6 | 62.4 | 1613 | US-09-061-154-1 | Sequence 1, Appli |

| | | | | | | | |
|---|-----|------|------|-------|---|----------------------|--------------------|
| C | 28 | 15.6 | 62.4 | 1613 | 4 | US-09-061-154-3 | Sequence 3, Appli |
| | 29 | 15.6 | 62.4 | 1654 | 3 | US-08-991-426-3 | Sequence 3, Appli |
| | 30 | 15.6 | 62.4 | 1654 | 3 | US-09-143-470-3 | Sequence 3, Appli |
| | 31 | 15.6 | 62.4 | 1654 | 4 | US-08-649-006A-3 | Sequence 3, Appli |
| | 32 | 15.6 | 62.4 | 1654 | 4 | US-09-771-023-3 | Sequence 3, Appli |
| | 33 | 15.6 | 62.4 | 2153 | 2 | US-08-864-804-2 | Sequence 2, Appli |
| | 34 | 15.6 | 62.4 | 2254 | 4 | US-09-138-277C-2 | Sequence 2, Appli |
| | 35 | 15.6 | 62.4 | 2447 | 1 | US-08-050-132A-1 | Sequence 1, Appli |
| | 36 | 15.6 | 62.4 | 2447 | 3 | US-08-750-222A-1 | Sequence 1, Appli |
| | 37 | 15.6 | 62.4 | 2447 | 3 | US-08-815-652B-1 | Sequence 1, Appli |
| | 38 | 15.6 | 62.4 | 2447 | 3 | US-08-254-353A-1 | Sequence 1, Appli |
| | 39 | 15.6 | 62.4 | 2447 | 5 | PCT-US92-05374A-1 | Sequence 1, Appli |
| | 40 | 15.6 | 62.4 | 2447 | 5 | PCT-US95-07084-1 | Sequence 1, Appli |
| | 41 | 15.6 | 62.4 | 3133 | 1 | US-08-162-809-1 | Sequence 1, Appli |
| | 42 | 15.6 | 62.4 | 4079 | 4 | US-09-016-434-1173 | Sequence 1173, Ap |
| | 43 | 15.6 | 62.4 | 4426 | 4 | US-09-658-687A-3 | Sequence 3, Appli |
| | 44 | 15.6 | 62.4 | 45546 | 4 | US-09-146-053-6 | Sequence 6, Appli |
| | 45 | 15.4 | 61.6 | 333 | 4 | US-09-489-039A-3453 | Sequence 3453, Ap |
| | 46 | 15.4 | 61.6 | 345 | 4 | US-09-127-946-5 | Sequence 5, Appli |
| | 47 | 15.4 | 61.6 | 390 | 4 | US-09-595-549-11 | Sequence 11, Appl |
| | 48 | 15.4 | 61.6 | 531 | 4 | US-09-976-594-436 | Sequence 436, App |
| | 49 | 15.4 | 61.6 | 555 | 4 | US-09-252-991A-5563 | Sequence 5563, App |
| | 50 | 15.4 | 61.6 | 807 | 4 | US-09-833-381-42 | Sequence 42, Appl |
| | 51 | 15.4 | 61.6 | 1098 | 4 | US-09-170-486D-117 | Sequence 117, App |
| | 52 | 15.4 | 61.6 | 1098 | 4 | US-09-170-486D-225 | Sequence 225, App |
| | 53 | 15.4 | 61.6 | 1280 | 3 | US-09-199-737-1 | Sequence 1, Appli |
| | 54 | 15.4 | 61.6 | 1280 | 4 | US-08-900-230-1 | Sequence 1, Appli |
| | 55 | 15.4 | 61.6 | 1280 | 4 | US-09-058-333A-1 | Sequence 1, Appli |
| | 56 | 15.4 | 61.6 | 1299 | 4 | US-09-252-991A-5631 | Sequence 5631, Ap |
| | 57 | 15.4 | 61.6 | 1326 | 4 | US-09-252-991A-15735 | Sequence 15735, A |
| | 58 | 15.4 | 61.6 | 1330 | 3 | US-08-888-077A-26 | Sequence 26, Appl |
| | 59 | 15.4 | 61.6 | 1330 | 4 | US-09-665-479A-15 | Sequence 15, Appl |
| | 60 | 15.4 | 61.6 | 1515 | 4 | US-09-252-991A-15701 | Sequence 15701, A |
| | 61 | 15.4 | 61.6 | 1597 | 2 | US-08-724-974A-1 | Sequence 1, Appli |
| | 62 | 15.4 | 61.6 | 1611 | 4 | US-09-252-991A-5600 | Sequence 5600, Ap |
| | 63 | 15.4 | 61.6 | 1697 | 4 | US-09-364-425B-26 | Sequence 26, Appl |
| | 64 | 15.4 | 61.6 | 1698 | 4 | US-09-489-039A-1732 | Sequence 1732, Ap |
| | 65 | 15.4 | 61.6 | 1794 | 3 | US-08-920-270-1 | Sequence 1, Appli |
| | 66 | 15.4 | 61.6 | 1866 | 3 | US-08-903-742-1 | Sequence 1, Appli |
| | 67 | 15.4 | 61.6 | 1866 | 3 | US-09-412-289-1 | Sequence 1, Appli |
| | 68 | 15.4 | 61.6 | 1981 | 4 | US-09-620-312D-870 | Sequence 870, App |
| | 69 | 15.4 | 61.6 | 2118 | 4 | US-09-252-991A-5605 | Sequence 5605, Ap |
| | 70 | 15.4 | 61.6 | 2614 | 2 | US-08-795-868-15 | Sequence 15, Appl |
| | 71 | 15.4 | 61.6 | 2614 | 4 | US-09-303-089-15 | Sequence 15, Appl |
| | 72 | 15.4 | 61.6 | 2614 | 4 | US-09-134-250-15 | Sequence 15, Appl |
| | 73 | 15.4 | 61.6 | 2874 | 4 | US-09-489-039A-1500 | Sequence 1500, Ap |
| | 74 | 15.4 | 61.6 | 2898 | 4 | US-09-051-696-7 | Sequence 7, Appli |
| | 75 | 15.4 | 61.6 | 3901 | 1 | US-08-326-297-1 | Sequence 1, Appli |
| | 76 | 15.4 | 61.6 | 3901 | 3 | US-08-617-454-1 | Sequence 1, Appli |
| | 77 | 15.4 | 61.6 | 3901 | 5 | PCT-US94-01144-1 | Sequence 1, Appli |
| | 78 | 15.4 | 61.6 | 4616 | 1 | US-08-340-203A-1 | Sequence 1, Appli |
| | 79 | 15.4 | 61.6 | 4616 | 2 | US-08-452-567-1 | Sequence 1, Appli |
| | 80 | 15.4 | 61.6 | 4616 | 2 | US-08-452-567-1 | Sequence 1, Appli |
| | 81 | 15.4 | 61.6 | 4616 | 3 | US-09-085-407-1 | Sequence 1, Appli |
| | 82 | 15.4 | 61.6 | 6420 | 2 | US-08-374-483-1 | Sequence 1, Appli |
| | 83 | 15.4 | 61.6 | 6803 | 3 | US-08-665-259-19 | Sequence 19, Appl |
| | 84 | 15.4 | 61.6 | 6803 | 3 | US-08-762-500-19 | Sequence 19, Appl |
| | 85 | 15.4 | 61.6 | 8236 | 1 | US-08-461-837-1 | Sequence 1, Appli |
| | 86 | 15.4 | 61.6 | 8236 | 3 | US-08-973-223-1 | Sequence 1, Appli |
| | 87 | 15.4 | 61.6 | 8236 | 3 | US-09-347-060-1 | Sequence 1, Appli |
| | 88 | 15.4 | 61.6 | 8236 | 5 | PCT-US96-09495-1 | Sequence 1, Appli |
| | 89 | 15.4 | 61.6 | 9592 | 1 | US-08-393-734-3 | Sequence 3, Appli |
| | 90 | 15.4 | 61.6 | 9592 | 3 | US-08-894-489-3 | Sequence 3, Appli |
| | 91 | 15.4 | 61.6 | 9641 | 2 | US-08-374-483-3 | Sequence 1, Appli |
| | 92 | 15.4 | 61.6 | 10398 | 2 | US-08-331-384-1 | Sequence 1, Appli |
| | 93 | 15.4 | 61.6 | 10398 | 2 | US-08-708-188-1 | Sequence 1, Appli |
| | 94 | 15.4 | 61.6 | 10398 | 2 | US-08-836-087-1 | Sequence 1, Appli |
| | 95 | 15.4 | 61.6 | 10398 | 3 | US-09-246-320-1 | Sequence 1, Appli |
| | 96 | 15.4 | 61.6 | 10398 | 3 | US-09-243-743-1 | Sequence 1, Appli |
| | 97 | 15.4 | 61.6 | 10398 | 4 | US-09-546-738-1 | Sequence 1, Appli |
| | 98 | 15.4 | 61.6 | 10398 | 4 | US-09-923-726-1 | Sequence 1, Appli |
| | 99 | 15.4 | 61.6 | 14672 | 4 | US-08-961-527-111 | Sequence 111, App |
| | 100 | 15.4 | 61.6 | 17000 | 4 | US-09-548-797B-7 | Sequence 7, Appli |

| | | | | | | | | | | | | | |
|-------|------|------|--------|---|---------------------|--------------------|-----|------|------|--------|---|---------------------|--------------------|
| C 101 | 15.4 | 61.6 | 32798 | 4 | US-09-604-694B-1 | Sequence 1, Appli | 174 | 15 | 60.0 | 5086 | 5 | PCT-US93-05851-4 | Sequence 4, Appli |
| C 102 | 15.4 | 61.6 | 34303 | 2 | US-08-735-609-4 | Sequence 4, Appli | 175 | 15 | 60.0 | 5086 | 5 | PCT-US93-06251-2 | Sequence 2, Appli |
| C 103 | 15.4 | 61.6 | 34303 | 2 | US-08-735-609-4 | Sequence 4, Appli | 176 | 15 | 60.0 | 5094 | 4 | US-09-234-186-7 | Sequence 7, Appli |
| C 104 | 15.4 | 61.6 | 34303 | 3 | US-09-315-372-4 | Sequence 4, Appli | 177 | 15 | 60.0 | 5104 | 6 | 5506344-1 | Patent No. 5506344 |
| C 105 | 15.4 | 61.6 | 34303 | 3 | US-09-244-752-4 | Sequence 4, Appli | 178 | 15 | 60.0 | 6030 | 4 | US-09-023-655-1015 | Sequence 1015, Ap |
| C 106 | 15.4 | 61.6 | 34303 | 3 | US-09-245-497-4 | Sequence 4, Appli | 179 | 15 | 60.0 | 44848 | 4 | US-09-435-739-42 | Sequence 42, Appli |
| C 107 | 15.4 | 61.6 | 34303 | 4 | US-09-562-919-4 | Sequence 4, Appli | 180 | 15 | 60.0 | 536165 | 4 | US-09-214-808-1 | Sequence 1, Appli |
| C 108 | 15.4 | 61.6 | 34382 | 2 | US-08-374-483-6 | Sequence 6, Appli | 181 | 15 | 60.0 | 786431 | 4 | US-09-751-389-3 | Sequence 3, Appli |
| C 109 | 15.4 | 61.6 | 35408 | 3 | US-08-973-334-3 | Sequence 3, Appli | 182 | 14.8 | 59.2 | 283 | 4 | US-09-313-294A-4190 | Sequence 4190, Ap |
| C 110 | 15.4 | 61.6 | 35408 | 3 | US-09-563-859A-3 | Sequence 3, Appli | 183 | 14.8 | 59.2 | 354 | 2 | US-08-908-713-3 | Sequence 3, Appli |
| C 111 | 15.4 | 61.6 | 35408 | 3 | US-08-549-489-3 | Sequence 3, Appli | 184 | 14.8 | 59.2 | 1150 | 3 | US-09-372-934-3 | Sequence 3, Appli |
| C 112 | 15.4 | 61.6 | 35871 | 4 | US-09-956-335-2 | Sequence 3, Appli | 185 | 14.8 | 59.2 | 1150 | 3 | US-09-766-916-3 | Sequence 3, Appli |
| C 113 | 15.4 | 61.6 | 35935 | 2 | US-08-735-609-1 | Sequence 1, Appli | 186 | 14.8 | 59.2 | 1150 | 4 | US-09-766-898-3 | Sequence 3, Appli |
| C 114 | 15.4 | 61.6 | 35935 | 2 | US-08-735-609-1 | Sequence 1, Appli | 187 | 14.8 | 59.2 | 1150 | 4 | US-09-637-040C-3 | Sequence 3, Appli |
| C 115 | 15.4 | 61.6 | 35935 | 3 | US-08-379-452-43 | Sequence 43, Appli | 188 | 14.8 | 59.2 | 1185 | 4 | US-09-134-001C-2117 | Sequence 2117, Ap |
| C 116 | 15.4 | 61.6 | 35935 | 3 | US-09-315-372-1 | Sequence 1, Appli | 189 | 14.8 | 59.2 | 1193 | 4 | US-09-227-357-63 | Sequence 63, Appli |
| C 117 | 15.4 | 61.6 | 35935 | 3 | US-09-244-752-1 | Sequence 1, Appli | 190 | 14.8 | 59.2 | 1301 | 3 | US-08-804-166-7 | Sequence 7, Appli |
| C 118 | 15.4 | 61.6 | 35935 | 3 | US-09-245-497-1 | Sequence 1, Appli | 191 | 14.8 | 59.2 | 1301 | 3 | US-08-910-991-7 | Sequence 7, Appli |
| C 119 | 15.4 | 61.6 | 35935 | 3 | US-09-409-670-43 | Sequence 43, Appli | 192 | 14.8 | 59.2 | 1301 | 3 | US-09-756-186-7 | Sequence 7, Appli |
| C 120 | 15.4 | 61.6 | 35935 | 4 | US-09-562-919-1 | Sequence 1, Appli | 193 | 14.8 | 59.2 | 1373 | 4 | US-09-482-273-70 | Sequence 70, Appli |
| C 121 | 15.4 | 61.6 | 35978 | 4 | US-09-956-335-1 | Sequence 1, Appli | 194 | 14.8 | 59.2 | 1526 | 3 | US-09-324-867-63 | Sequence 63, Appli |
| C 122 | 15.4 | 61.6 | 786431 | 1 | US-09-751-389-3 | Sequence 3, Appli | 195 | 14.8 | 59.2 | 1717 | 1 | US-08-229-515A-12 | Sequence 12, Appli |
| C 123 | 15.2 | 60.8 | 48 | 1 | US-08-225-224-26 | Sequence 26, Appli | 196 | 14.8 | 59.2 | 1717 | 1 | US-08-645-865-12 | Sequence 12, Appli |
| C 124 | 15.2 | 60.8 | 48 | 3 | US-08-722-258-26 | Sequence 26, Appli | 197 | 14.8 | 59.2 | 2230 | 4 | US-09-189-527-12 | Sequence 12, Appli |
| C 125 | 15.2 | 60.8 | 48 | 5 | PCT-US93-04868-26 | Sequence 26, Appli | 198 | 14.8 | 59.2 | 2508 | 4 | US-09-758-282B-48 | Sequence 48, Appli |
| C 126 | 15.2 | 60.8 | 134 | 3 | US-09-284-782-27 | Sequence 27, Appli | 199 | 14.8 | 59.2 | 2508 | 4 | US-09-758-282B-242 | Sequence 242, App |
| C 127 | 15.2 | 60.8 | 294 | 4 | US-09-489-039A-3664 | Sequence 3664, Ap | 200 | 14.8 | 59.2 | 2526 | 4 | US-09-758-282B-258 | Sequence 258, App |
| C 128 | 15.2 | 60.8 | 1080 | 2 | US-08-918-727-2 | Sequence 2, Appli | 201 | 14.8 | 59.2 | 2526 | 4 | US-09-758-282B-262 | Sequence 262, App |
| C 129 | 15.2 | 60.8 | 1080 | 3 | US-09-205-680A-2 | Sequence 2, Appli | 202 | 14.8 | 59.2 | 2831 | 2 | US-08-906-713-1 | Sequence 1, Appli |
| C 130 | 15.2 | 60.8 | 1275 | 4 | US-08-614-155B-4 | Sequence 4, Appli | 203 | 14.8 | 59.2 | 3158 | 2 | US-08-464-517-36 | Sequence 36, Appli |
| C 131 | 15.2 | 60.8 | 1275 | 4 | US-09-166-963-4 | Sequence 4, Appli | 204 | 14.8 | 59.2 | 3158 | 2 | US-08-245-351A-36 | Sequence 36, Appli |
| C 132 | 15.2 | 60.8 | 1630 | 3 | US-09-383-586-4 | Sequence 4, Appli | 205 | 14.8 | 59.2 | 3158 | 3 | US-08-463-712-36 | Sequence 36, Appli |
| C 133 | 15.2 | 60.8 | 1698 | 4 | US-09-489-039A-3860 | Sequence 3860, Ap | 206 | 14.8 | 59.2 | 3394 | 4 | US-09-620-312D-314 | Sequence 314, App |
| C 134 | 15.2 | 60.8 | 2160 | 4 | US-09-372-668-1 | Sequence 1, Appli | 207 | 14.8 | 59.2 | 5207 | 4 | US-09-858-664A-1 | Sequence 1, Appli |
| C 135 | 15.2 | 60.8 | 2160 | 4 | US-09-697-377-1 | Sequence 1, Appli | 208 | 14.8 | 59.2 | 5207 | 4 | US-10-274-978-1 | Sequence 1, Appli |
| C 136 | 15.2 | 60.8 | 2160 | 4 | US-09-696-868-1 | Sequence 1, Appli | 209 | 14.8 | 59.2 | 5207 | 4 | US-10-274-978-3 | Sequence 3, Appli |
| C 137 | 15.2 | 60.8 | 2160 | 4 | US-09-697-341-1 | Sequence 1, Appli | 210 | 14.8 | 59.2 | 5683 | 4 | US-09-845-583A-7 | Sequence 7, Appli |
| C 138 | 15.2 | 60.8 | 2448 | 3 | US-08-691-563C-53 | Sequence 53, Appli | 211 | 14.8 | 59.2 | 5702 | 4 | US-09-566-921-118 | Sequence 118, App |
| C 139 | 15.2 | 60.8 | 2448 | 3 | US-09-374-766-53 | Sequence 53, Appli | 212 | 14.8 | 59.2 | 7626 | 4 | US-09-576-594-14 | Sequence 14, Appli |
| C 140 | 15.2 | 60.8 | 2448 | 4 | US-08-979-847B-49 | Sequence 49, Appli | 213 | 14.8 | 59.2 | 35100 | 1 | US-08-306-991B-19 | Sequence 19, Appli |
| C 141 | 15.2 | 60.8 | 2556 | 4 | US-08-976-259-61 | Sequence 61, Appli | 214 | 14.8 | 59.2 | 35100 | 5 | PCT-US93-06251-19 | Sequence 19, Appli |
| C 142 | 15.2 | 60.8 | 2675 | 4 | US-09-636-791A-12 | Sequence 12, Appli | 215 | 14.8 | 59.2 | 75395 | 4 | US-09-984-890-3 | Sequence 3, Appli |
| C 143 | 15.2 | 60.8 | 4550 | 4 | US-09-462-136-1 | Sequence 1, Appli | 216 | 14.8 | 58.4 | 23 | 4 | US-09-710-200-29 | Sequence 29, Appli |
| C 144 | 15.2 | 60.8 | 5474 | 3 | US-09-040-738-1 | Sequence 1, Appli | 217 | 14.6 | 58.4 | 23 | 4 | US-09-975-408-29 | Sequence 29, Appli |
| C 145 | 15.2 | 60.8 | 5474 | 3 | US-08-652-426A-1 | Sequence 1, Appli | 218 | 14.6 | 58.4 | 37 | 3 | US-09-158-863C-33 | Sequence 33, Appli |
| C 146 | 15.2 | 60.8 | 2451 | 3 | US-08-652-426A-1 | Sequence 1, Appli | 219 | 14.6 | 58.4 | 80 | 2 | US-07-733-095B-10 | Sequence 10, Appli |
| C 147 | 15.2 | 60.8 | 59065 | 4 | US-09-433-702B-173 | Sequence 173, App | 220 | 14.6 | 58.4 | 120 | 6 | 5510472-5 | Patent No. 5510472 |
| C 148 | 15.2 | 60.8 | 59065 | 4 | US-09-978-197-3 | Sequence 3, Appli | 221 | 14.6 | 58.4 | 197 | 4 | US-09-189-618A-15 | Sequence 15, Appli |
| C 149 | 15.2 | 60.8 | 193303 | 4 | US-09-497-855A-37 | Sequence 37, Appli | 222 | 14.6 | 58.4 | 202 | 4 | US-09-313-294A-5307 | Sequence 5307, Ap |
| C 150 | 15.2 | 60.8 | 193303 | 4 | US-09-497-855A-44 | Sequence 44, Appli | 223 | 14.6 | 58.4 | 297 | 4 | US-09-313-294A-6906 | Sequence 6906, Ap |
| C 151 | 15 | 60.0 | 47 | 4 | US-09-497-855A-44 | Sequence 44, Appli | 224 | 14.6 | 58.4 | 304 | 4 | US-09-621-976-10937 | Sequence 10937, A |
| C 152 | 15 | 60.0 | 297 | 4 | US-08-422-978-2989 | Sequence 2989, Ap | 225 | 14.6 | 58.4 | 344 | 4 | US-09-621-976-10937 | Sequence 10937, A |
| C 153 | 15 | 60.0 | 409 | 4 | US-08-313-294A-381 | Sequence 31, Appli | 226 | 14.6 | 58.4 | 351 | 4 | US-09-433-313-361 | Sequence 361, App |
| C 154 | 15 | 60.0 | 431 | 4 | US-08-312-283C-356 | Sequence 356, App | 227 | 14.6 | 58.4 | 351 | 4 | US-09-352-616A-361 | Sequence 361, App |
| C 155 | 15 | 60.0 | 507 | 4 | US-09-105-470B-28 | Sequence 28, Appli | 228 | 14.6 | 58.4 | 351 | 4 | US-09-636-215-361 | Sequence 361, App |
| C 156 | 15 | 60.0 | 525 | 4 | US-09-621-976-18706 | Sequence 18706, A | 229 | 14.6 | 58.4 | 351 | 4 | US-09-685-166A-361 | Sequence 361, App |
| C 157 | 15 | 60.0 | 753 | 4 | US-09-489-039A-210 | Sequence 210, App | 230 | 14.6 | 58.4 | 365 | 4 | US-09-423-233-5 | Sequence 5, Appli |
| C 158 | 15 | 60.0 | 1050 | 4 | US-09-489-039A-968 | Sequence 968, App | 231 | 14.6 | 58.4 | 378 | 4 | US-09-621-976-18899 | Sequence 18899, A |
| C 159 | 15 | 60.0 | 1075 | 2 | US-08-809-960-1 | Sequence 1, Appli | 232 | 14.6 | 58.4 | 404 | 6 | 5510472-4 | Patent No. 5510472 |
| C 160 | 15 | 60.0 | 1410 | 2 | US-08-975-316-86 | Sequence 86, Appli | 233 | 14.6 | 58.4 | 407 | 3 | US-08-444-818-90 | Sequence 90, Appli |
| C 161 | 15 | 60.0 | 1410 | 4 | US-09-635-192A-86 | Sequence 86, Appli | 234 | 14.6 | 58.4 | 412 | 4 | US-09-621-976-18898 | Sequence 18898, A |
| C 162 | 15 | 60.0 | 1410 | 4 | US-09-169-789-86 | Sequence 86, Appli | 235 | 14.6 | 58.4 | 415 | 4 | US-09-635-747-39 | Sequence 39, Appli |
| C 163 | 15 | 60.0 | 2172 | 4 | US-09-484-970B-126 | Sequence 126, App | 236 | 14.6 | 58.4 | 415 | 4 | US-09-635-747-40 | Sequence 40, Appli |
| C 164 | 15 | 60.0 | 2461 | 1 | US-08-282-141-1 | Sequence 1, Appli | 237 | 14.6 | 58.4 | 437 | 4 | US-09-397-787-332 | Sequence 332, App |
| C 165 | 15 | 60.0 | 2555 | 4 | US-09-866-028-68 | Sequence 68, Appli | 238 | 14.6 | 58.4 | 441 | 4 | US-09-389-681-186 | Sequence 186, App |
| C 166 | 15 | 60.0 | 2852 | 3 | US-09-063-950-1 | Sequence 1, Appli | 239 | 14.6 | 58.4 | 441 | 4 | US-09-620-405B-186 | Sequence 186, App |
| C 167 | 15 | 60.0 | 4825 | 6 | 5459251-1 | Patent No. 5459251 | 240 | 14.6 | 58.4 | 441 | 4 | US-09-339-338-186 | Sequence 186, App |
| C 168 | 15 | 60.0 | 5086 | 2 | US-08-485-485A-19 | Sequence 19, Appli | 241 | 14.6 | 58.4 | 441 | 4 | US-09-433-826B-186 | Sequence 186, App |
| C 169 | 15 | 60.0 | 5086 | 2 | US-09-365-486A-14 | Sequence 14, Appli | 242 | 14.6 | 58.4 | 441 | 4 | US-09-604-287A-186 | Sequence 186, App |
| C 170 | 15 | 60.0 | 5086 | 3 | US-09-080-285-19 | Sequence 19, Appli | 243 | 14.6 | 58.4 | 441 | 4 | US-09-489-039A-6675 | Sequence 6675, Ap |
| C 171 | 15 | 60.0 | 5086 | 3 | US-08-880-342-14 | Sequence 14, Appli | 244 | 14.6 | 58.4 | 441 | 4 | US-09-834-759-186 | Sequence 186, App |
| C 172 | 15 | 60.0 | 5086 | 4 | US-09-724-426-19 | Sequence 19, Appli | 245 | 14.6 | 58.4 | 459 | 4 | US-09-489-039A-4811 | Sequence 4811, Ap |
| C 173 | 15 | 60.0 | 5086 | 4 | US-09-233-527-7 | Sequence 7, Appli | 246 | 14.6 | 58.4 | 461 | 4 | US-09-620-312D-244 | Sequence 244, App |

| | | | | | |
|------|------|-----|---|--------------------|---------------------|
| 14.6 | 58.4 | 473 | 4 | US-09-740-235-31 | Sequence 31, Appl |
| 14.6 | 58.4 | 501 | 2 | US-08-483-695-27 | Sequence 27, Appl |
| 14.6 | 58.4 | 501 | 2 | US-08-483-695-29 | Sequence 29, Appl |
| 14.6 | 58.4 | 501 | 2 | US-07-965-285-27 | Sequence 27, Appl |
| 14.6 | 58.4 | 501 | 2 | US-07-965-285-29 | Sequence 29, Appl |
| 14.6 | 58.4 | 501 | 2 | US-08-487-231-27 | Sequence 27, Appl |
| 14.6 | 58.4 | 501 | 2 | US-08-487-231-29 | Sequence 29, Appl |
| 14.6 | 58.4 | 501 | 2 | US-08-201-912-27 | Sequence 27, Appl |
| 14.6 | 58.4 | 501 | 3 | US-08-201-912-29 | Sequence 29, Appl |
| 14.6 | 58.4 | 501 | 3 | US-08-201-912-29 | Sequence 29, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-971-52 | Sequence 52, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-971-53 | Sequence 53, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-971-54 | Sequence 54, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-971-55 | Sequence 55, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-971-56 | Sequence 56, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-971-57 | Sequence 57, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-221-653-52 | Sequence 52, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-221-653-53 | Sequence 53, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-221-653-54 | Sequence 54, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-221-653-55 | Sequence 55, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-221-653-56 | Sequence 56, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-221-653-57 | Sequence 57, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-442-1443-52 | Sequence 52, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-442-1443-53 | Sequence 53, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-442-1443-54 | Sequence 54, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-442-1443-55 | Sequence 55, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-442-1443-56 | Sequence 56, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-442-1443-57 | Sequence 57, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-970-52 | Sequence 52, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-970-53 | Sequence 53, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-970-54 | Sequence 54, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-970-55 | Sequence 55, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-970-56 | Sequence 56, Appl |
| 14.6 | 58.4 | 549 | 3 | US-08-441-970-57 | Sequence 57, Appl |
| 14.6 | 58.4 | 550 | 4 | US-03-833-381-1388 | Sequence 1388, Appl |
| 14.6 | 58.4 | 561 | 4 | US-03-833-381-2014 | Sequence 2014, Appl |
| 14.6 | 58.4 | 573 | 2 | US-08-230-665A-104 | Sequence 104, Appl |
| 14.6 | 58.4 | 573 | 2 | US-08-230-665A-106 | Sequence 106, Appl |
| 14.6 | 58.4 | 573 | 2 | US-08-230-665A-107 | Sequence 107, Appl |
| 14.6 | 58.4 | 573 | 3 | US-08-078-271B-2 | Sequence 2, Appl |
| 14.6 | 58.4 | 573 | 3 | US-08-078-271B-3 | Sequence 3, Appl |
| 14.6 | 58.4 | 573 | 5 | PCT-US95-10398-104 | Sequence 104, Appl |
| 14.6 | 58.4 | 573 | 5 | PCT-US95-10398-106 | Sequence 106, Appl |
| 14.6 | 58.4 | 573 | 5 | PCT-US95-10398-107 | Sequence 107, Appl |
| 14.6 | 58.4 | 585 | 3 | US-08-444-818-161 | Sequence 161, Appl |
| 14.6 | 58.4 | 626 | 1 | US-08-233-608-3 | Sequence 3, Appl |
| 14.6 | 58.4 | 626 | 2 | US-08-887-480-3 | Sequence 3, Appl |
| 14.6 | 58.4 | 626 | 4 | US-09-635-747-38 | Sequence 38, Appl |
| 14.6 | 58.4 | 626 | 5 | PCT-US95-04712-3 | Sequence 3, Appl |
| 14.6 | 58.4 | 627 | 1 | US-08-233-608-4 | Sequence 4, Appl |
| 14.6 | 58.4 | 627 | 1 | US-08-233-608-47 | Sequence 47, Appl |
| 14.6 | 58.4 | 627 | 1 | US-08-887-480-4 | Sequence 47, Appl |
| 14.6 | 58.4 | 627 | 2 | US-08-732-187-4 | Sequence 47, Appl |
| 14.6 | 58.4 | 627 | 2 | US-08-732-187-47 | Sequence 47, Appl |

RESULT 1

```

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; APPLICANT: Verema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 16595
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
US-09-146-053-7
Query Match      68.8%; Score 17.2; DB 4; Length 16595;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGGCACATAGC 25
Db 12850 GGAGGGCCACAGGGCCCAAGC 12871

RESULT 4
US-09-621-976-167/c
Sequence 167, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
FILE OF INVENTION: ESTs and Encoded Human Proteins.
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 167
LENGTH: 481
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 297..479
NAME/KEY: sig_peptide
LOCATION: 297..353
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.4000000536743
OTHER INFORMATION: seq LCCSGCVPSLCCS/SY
US-09-621-976-167

Query Match      67.2%; Score 16.8; DB 4; Length 481;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGGCACATAGC 25
Db 376 GGAGCAACAGAGGGSACATAGC 355

RESULT 5
US-09-604-978-5/c
Sequence 5, Application US/09604978
Patent No. 6455674
GENERAL INFORMATION:
APPLICANT: Einat, Paz
APPLICANT: Skaliter, Rami
TITLE OF INVENTION: HYPOXIA-REGULATED GENES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: KOHN & ASSOCIATES
STREET: 30500 No. 6455674thwestern Hwy., Suite 401
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/604,978
FILING DATE: 28-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/138,112
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 0168.00034
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 5395055
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2059 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-09-146-053-7
Query Match      68.8%; Score 17.2; DB 4; Length 16595;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGGCACATAGC 25
Db 12850 GGAGGGCCACAGGGCCCAAGC 12871

RESULT 4
US-09-621-976-167/c
Sequence 167, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
FILE OF INVENTION: ESTs and Encoded Human Proteins.
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 167
LENGTH: 481
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 297..479
NAME/KEY: sig_peptide
LOCATION: 297..353
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.4000000536743
OTHER INFORMATION: seq LCCSGCVPSLCCS/SY
US-09-621-976-167

Query Match      67.2%; Score 16.8; DB 4; Length 481;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGGCACATAGC 25
Db 376 GGAGCAACAGAGGGSACATAGC 355

RESULT 5
US-09-604-978-5/c
Sequence 5, Application US/09604978
Patent No. 6455674
GENERAL INFORMATION:
APPLICANT: Einat, Paz
APPLICANT: Skaliter, Rami
TITLE OF INVENTION: HYPOXIA-REGULATED GENES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: KOHN & ASSOCIATES
STREET: 30500 No. 6455674thwestern Hwy., Suite 401
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/604,978
FILING DATE: 28-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/138,112
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 0168.00034
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 5395055
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2059 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-09-604-978-5
Query Match      67.2%; Score 16.8; DB 4; Length 2059;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 AGGGCCAGAGGGGCACATAGC 25
Db 154 AGGGCCAGAGGGGCACATAGC 135

RESULT 6
US-09-604-728-5/c
Sequence 5, Application US/09604728
Patent No. 6555667
GENERAL INFORMATION:
APPLICANT: Einat, Paz
APPLICANT: Skaliter, Rami
TITLE OF INVENTION: HYPOXIA-REGULATED GENES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: KOHN & ASSOCIATES
STREET: 30500 No. 6555667thwestern Hwy., Suite 401
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/604,728
FILING DATE: 28-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/138,112
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 0168.00034
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 5395055
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2059 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
```


US-09-604-728-5
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 5;
Query Match 67.2%; Score 16.8; DB 4; Length 2059;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 AGGCCAGGCGACATAGC 25
Db 154 AGGCCAGGCGACATAGC 135

RESULT 7
US-08-049-283A-1/c
; Sequence 1, Application US/08049283A
; Patent No. 5502176
; GENERAL INFORMATION:
; APPLICANT: Tenen, Daniel G.
; APPLICANT: Pahl, Heike L.
; APPLICANT: Burn, Timothy C.
; TITLE OF INVENTION: Cell Specific Promoter and Uses Thereof
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,283A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/020,465
; FILING DATE: 19-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/837,776
; FILING DATE: 13-FEB-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: BIH91-03/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-049-283A-1

Query Match 66.4%; Score 16.6; DB 1; Length 533;
Best Local Similarity 82.6%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CGGAGGCGACATAGC 25
Db 351 CGGAGGCGACATAGC 329

RESULT 8
US-08-618-100B-3
; Sequence 3, Application US/08618100B
; Patent No. 6068976
; GENERAL INFORMATION:
; APPLICANT: Briggs, Michael R.
; APPLICANT: Auwerx, Johan
; APPLICANT: de Vos, Piet
; APPLICANT: Staels, Bart
; APPLICANT: Croston, Glenn E.
; APPLICANT: Miller, Stephen G.
; TITLE OF INVENTION: MODULATORS OF ob GENE AND
; TITLE OF INVENTION: SCREENING METHODS THEREFOR
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,100B
; FILING DATE: March 19, 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/558,588
; FILING DATE: October 30, 1995
; APPLICATION NUMBER: 08/510,584
; FILING DATE: August 2, 1995
; APPLICATION NUMBER: 08/418,096
; FILING DATE: April 5, 1995
; APPLICATION NUMBER: 08/408,584
; FILING DATE: March 20, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 219/075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10684 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Sequence between exon 1 and exon 2
; Patent No. 6068976
US-08-618-100B-3

Query Match 66.4%; Score 16.6; DB 3; Length 10684;
Best Local Similarity 82.6%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCGAGGCGACATAGC 24
Db 2252 CCGGCTAGAGGCGACATAG 2274

RESULT 9
US-09-277-262-3/c
; Sequence 3, Application US/09277262
; Patent No. 6395482

/ GENERAL INFORMATION:
/ APPLICANT: Karayiorgou, Maria
/ APPLICANT: Gogos, Joseph A
/ TITLE OF INVENTION: METHODS OF DETERMINING A SUSCEPTIBILITY TO OR PRESENCE
/ TITLE OF INVENTION: OF SCHIZOPHRENIA, OR A DISEASE OR DISORDER RELATED
/ TITLE OF INVENTION: THERE TO
/ FILE REFERENCE: 600-1-223 CIP
/ CURRENT APPLICATION NUMBER: US/09/277,262
/ CURRENT FILING DATE: 1999-03-26
/ EARLIER APPLICATION NUMBER: 09/229,530
/ EARLIER FILING DATE: 1999-01-13
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 3
/ LENGTH: 1494
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ US-09-277-262-3

Query Match 65.6%; Score 16.4; DB 4; Length 1494;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AGGGCCAGGCGCACATA 23
DB 1350 AGGGCCATAGGGCCACATA 1333

RESULT 10
US-09-277-262-7/c
/ Sequence 7, Application US/09277262
/ Patent No. 6395482
/ GENERAL INFORMATION:
/ APPLICANT: Karayiorgou, Maria
/ APPLICANT: Gogos, Joseph A
/ TITLE OF INVENTION: METHODS OF DETERMINING A SUSCEPTIBILITY TO OR PRESENCE
/ TITLE OF INVENTION: OF SCHIZOPHRENIA, OR A DISEASE OR DISORDER RELATED
/ TITLE OF INVENTION: THERE TO
/ FILE REFERENCE: 600-1-223 CIP
/ CURRENT APPLICATION NUMBER: US/09/277,262
/ CURRENT FILING DATE: 1999-03-26
/ EARLIER APPLICATION NUMBER: 09/229,530
/ EARLIER FILING DATE: 1999-01-13
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 2240
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ US-09-277-262-7

Query Match 65.6%; Score 16.4; DB 4; Length 2240;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 AGGGCCAGGCGCACATA 23
DB 1646 AGGGCCATAGGGCCACATA 1629

RESULT 11
US-09-482-273-44
/ Sequence 44, Application US/09482273
/ Patent No. 6534631
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 71 Human Secreted Proteins
/ FILE REFERENCE: P2030P1
/ CURRENT APPLICATION NUMBER: US/09/482,273
/ CURRENT FILING DATE: 2000-01-13
/ EARLIER APPLICATION NUMBER: PCT/US99/15849
/ EARLIER FILING DATE: 1999-07-14
/ EARLIER APPLICATION NUMBER: 60/092,921

/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: 60/092,922
/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: 60/092,956
/ EARLIER FILING DATE: 1998-07-15
/ NUMBER OF SEQ ID NOS: 267
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 44
/ LENGTH: 1262
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-482-273-44

Query Match 64.8%; Score 16.2; DB 4; Length 1262;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGGCGCACAT 22
DB 1004 CCGAGGGCCAGGCGCACAT 1024

RESULT 12
US-09-448-176-9
/ Sequence 9, Application US/09448176
/ Patent No. 6248533
/ GENERAL INFORMATION:
/ APPLICANT: KAMIZONO, Shintaro
/ YAMADA, Akira
/ HIGUCHI, Takafumi
/ KATO, Hirohisa
/ ITOH, Kyogo
/ SEKI, Naoko
/ TITLE OF INVENTION: GENE DIAGNOSIS OF DISEASES WHEREIN
/ TNF-ALPHA PROMOTERS PARTICIPATE
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FOLEY & LARDNER
/ STREET: 3000 K Street, N.W.
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20007-5109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/448,176
/ FILING DATE: 24-No. 6248533-1999
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: WO PCT/JP98/02270
/ FILING DATE: 25-MAY-1998
/ APPLICATION NUMBER: WO PCT/JP97/04304
/ FILING DATE: 26-NOV-1997
/ APPLICATION NUMBER: JP 9-173900
/ FILING DATE: 30-JUN-1997
/ APPLICATION NUMBER: JP 9-134973
/ FILING DATE: 26-MAY-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Wegner, Harold C.
/ REGISTRATION NUMBER: 25,258
/ REFERENCE/DOCKET NUMBER: 74129/472
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 672-5300
/ TELEFAX: (202) 672-5399
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1357 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double

QY 4 GGAGGGCCAGAGGGCACATAG 24
DB 56 GGAGGGACAGAGGGGCTCAAAG 76

RESULT 14
US-09-448-176-11
; Sequence 11, Application US/09448176
; Patent No. 6248533
; GENERAL INFORMATION:
; APPLICANT: KAMIZONO, Shintaro
; YAMADA, Akira
; HIGUCHI, Takafumi
; KATO, Hirohisa
; ITOH, Kyogo
; SEKI, Naoko
; TITLE OF INVENTION: GENE DIAGNOSIS OF DISEASES WHEREIN
; TNF-ALPHA PROMOTERS PARTICIPATE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/448,176
; FILING DATE: 24-NOV-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP98/02270
; FILING DATE: 25-MAY-1998
; APPLICATION NUMBER: WO PCT/JP97/04304
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: JP 9-173900
; FILING DATE: 30-JUN-1997
; APPLICATION NUMBER: JP 9-134973
; FILING DATE: 26-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/472
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-448-176-11

Query Match 64.8%; Score 16.2; DB 3; Length 1357;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGCACATAG 24
DB 56 GGAGGGACAGAGGGGCTCAAAG 76

RESULT 15
US-08-762-428A-5/c

QY 4 GGAGGGCCAGAGGGCACATAG 24
DB 56 GGAGGGACAGAGGGGCTCAAAG 76

RESULT 13
US-09-448-176-10
; Sequence 10, Application US/09448176
; Patent No. 6248533
; GENERAL INFORMATION:
; APPLICANT: KAMIZONO, Shintaro
; YAMADA, Akira
; HIGUCHI, Takafumi
; KATO, Hirohisa
; ITOH, Kyogo
; SEKI, Naoko
; TITLE OF INVENTION: GENE DIAGNOSIS OF DISEASES WHEREIN
; TNF-ALPHA PROMOTERS PARTICIPATE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/448,176
; FILING DATE: 24-NOV-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP98/02270
; FILING DATE: 25-MAY-1998
; APPLICATION NUMBER: WO PCT/JP97/04304
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: JP 9-173900
; FILING DATE: 30-JUN-1997
; APPLICATION NUMBER: JP 9-134973
; FILING DATE: 26-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/472
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-448-176-10

Query Match 64.8%; Score 16.2; DB 3; Length 1357;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

/ Sequence 5, Application US/08762428A
/ Patent No. 6120993
/ GENERAL INFORMATION:
/ APPLICANT: Ye, Guo-jie
/ APPLICANT: Breslow, Esther M.
/ APPLICANT: Meister, Alton
/ TITLE OF INVENTION: 5-OXOPROLINASE
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
/ STREET: Clinton Square, P.O. Box 1051
/ CITY: Rochester
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 14603
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/762,428A
/ FILING DATE:
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: TIMIAN, SUSAN J.
/ REGISTRATION NUMBER: 34,103
/ REFERENCE/DOCKET NUMBER: 19603/1330
/ TELEPHONE: 716-263-1636
/ TELEFAX: 716-263-1600
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3867 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
/ US-08-762-428A-5

Query Match 64.8%; Score 16.2; DB 3; Length 3867;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGCACA 21
DB 875 CCAGAGGGCCAGAGGCACA 855

RESULT 16
US-08-762-428A-7/c
/ Sequence 7, Application US/08762428A
/ Patent No. 6120993
/ GENERAL INFORMATION:
/ APPLICANT: Ye, Guo-jie
/ APPLICANT: Breslow, Esther M.
/ APPLICANT: Meister, Alton
/ TITLE OF INVENTION: 5-OXOPROLINASE
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
/ STREET: Clinton Square, P.O. Box 1051
/ CITY: Rochester
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 14603
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/762,428A

/ FILING DATE:
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: TIMIAN, SUSAN J.
/ REGISTRATION NUMBER: 34,103
/ REFERENCE/DOCKET NUMBER: 19603/1330
/ TELEPHONE: 716-263-1636
/ TELEFAX: 716-263-1600
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4016 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
/ US-08-762-428A-7

Query Match 64.8%; Score 16.2; DB 3; Length 4016;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGCACA 21
DB 980 CCAGAGGGCCAGAGGCACA 960

RESULT 17
US-09-050-159-123/c
/ Sequence 123, Application US/09050159A
/ Patent No. 6197505
/ GENERAL INFORMATION:
/ APPLICANT: No. 6197505berg, Leif T
/ APPLICANT: Andersson, Maria K
/ APPLICANT: Linstrom, Per H
/ TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
/ TITLE OF INVENTION: COMPOSITIONS FOR USE THEREOF
/ FILE REFERENCE: 1248/1D042
/ CURRENT APPLICATION NUMBER: US/09/050,159A
/ CURRENT FILING DATE: 1998-03-27
/ EARLIER APPLICATION NUMBER: 60/042,930
/ EARLIER FILING DATE: 1987-04-03
/ NUMBER OF SEQ ID NOS: 133
/ SOFTWARE: Patent in Ver. 2.1
/ SEQ ID NO 123
/ LENGTH: 1278
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: Angiotensinogen, 5' region and exon 1
/ US-09-050-159-123

Query Match 64.0%; Score 16; DB 3; Length 1278;
Best Local Similarity 79.2%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CCGGAGGGCCAGAGGCACATAGC 25
DB 840 CCGGAGGTGCAGAGGGCCAGAGGGC 817

RESULT 18
US-09-976-594-193
/ Sequence 193, Application US/09976594
/ Patent No. 6673549
/ GENERAL INFORMATION:
/ APPLICANT: Furness, Michael
/ APPLICANT: Buchbinder, Jenny
/ TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
/ FILE REFERENCE: PA-0041 US
/ CURRENT APPLICATION NUMBER: US/09/976,594
/ CURRENT FILING DATE: 2001-10-12

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 193
; LENGTH: 1973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 235356.1
; NAME/KEY: unsure
; LOCATION: 614
; OTHER INFORMATION: a, t, c, g, or other
US-09-376-594-193

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 1973;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CCGGAGGCCAGAGGCCACATAG 25
Db 1437 CCTGAGGCCAGAGTGACACAGC 1460

RESULT 19
US-09-566-921-52/c
; Sequence 52 Application US/09566921
; Patent No. 6692888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 52
; LENGTH: 2201
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6692888 223035.8
US-09-566-921-52

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 2201;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CCGGAGGCCAGAGGCCACATAG 24
Db 111 CCCCTGGGCCAGAGGCCACATGG 88

RESULT 20
US-09-462-917A-106
; Sequence 106 Application US/09462917A
; Patent No. 6511801
; GENERAL INFORMATION:
; APPLICANT: Delaporte, Eric
; APPLICANT: Peeters, Martine
; APPLICANT: Saman, Eric
; APPLICANT: Vanden Haesevelde, Marlen
; TITLE OF INVENTION: HIV-1 Group O Antigens and Uses Thereof
; FILE REFERENCE: INNS:014 11362.0014 NPUS00
; CURRENT APPLICATION NUMBER: US/09/462,917A
; CURRENT FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: PCT/EP98/04522
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 106
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Human
US-09-462-917A-106

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 3600;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CCGGAGGCCAGAGGCCACATAG 24
Db 797 CCCAGAGACCAGGGGCCACAGAG 820

RESULT 21
US-08-943-731-4/c
; Sequence 4 Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DELTAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISA A.
; APPLICANT: LARSON, ANDREA W.
; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORKKO, JARMO
; APPLICANT: ALA-KOKKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,731
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9599-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17606 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-943-731-4

Query Match 64.0%; Score 16; DB 3; Length 17606;
Best Local Similarity 79.2%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels

Qy 1 CCGGAGGGCCAGGGCACATAG 24
||| ||| ||| ||| ||| ||| ||| |||
Db 15602 CCCGAGGGCCAGGGTGCCCATGG 15579

RESULT 22

```

US-09-015-188-5/C
; Sequence 5, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: "n" at all locations is "unknown"
US-09-015-188-5

```

```
Query Match      63.2%; Score 15.8; DB 4; Length 1231;
Best Local Similarity 77.3%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 5; Indels 0
```

QY
db

1 CCCGAGGGCCAGAGGCACAT 22
643 CCGGGNGGCCAGAGGGNACNT 622

RESULT 23

```

US-09-210-748A-3
; Sequence 3, Application US/09210748A
; Patent No. 6335156
; GENERAL INFORMATION:
; APPLICANT: Hermeking, Heiko
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: 14-3-3 SIGMA ARREST THE CELL CYCLE
; FILE REFERENCE: 1107.77810
; CURRENT APPLICATION NUMBER: US/09/210,748A
; CURRENT FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/069,416
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-210-748A-3

```

Query Match 63.2%; Score 15.8; DB 4; Length 7680;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0

Qy 4 GGAGGGCCAGAGGCACAT 22
db 7601 GGAGGGCCAGAGGGTAAT 7619

RESULT 24

US-08-764-233A-1/c

```

/ Sequence 1, Application US/09764233A
/ Patent No. 5716849
/ GENERAL INFORMATION:
/ APPLICANT: Ligon, James M.
/ APPLICANT: Schupp, Thomas
/ APPLICANT: Beck, James J.
/ APPLICANT: Hill, Dwight S.
/ APPLICANT: Neff, Snezana
/ APPLICANT: Ryals, John A.
/ TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Ciba-Geigy Corporation
/ STREET: 520 White Plains Road, P.O. Box 2005
/ CITY: Tarrytown
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10591
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0; Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/764,233A
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/729,214
/ FILING DATE: 09-OCT-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/258,261
/ FILING DATE: 08-JUN-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Meigs, J. Timothy
/ REGISTRATION NUMBER: 38,241
/ REFERENCE/DOCKET NUMBER: 1506/CIP6
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (919) 541-8587
/ TELEFAX: (919) 541-8689
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 49377 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: Sorangium cellulosum
/ CLONE: p98/1, pJL3, and pVKM15
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 383..760
/ OTHER INFORMATION: /product= "SorR"
/ OTHER INFORMATION: /note= "This gene encodes a protein th
/ OTHER INFORMATION: the reductase domains of type I PKSS s
/ OTHER INFORMATION: Saccharopolyspora erythraea."
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 927..19874
/ OTHER INFORMATION: /product= "SorA"
/ OTHER INFORMATION: /note= "Gene product is highly homolog
/ OTHER INFORMATION: are known to be involved in the synthe
/ OTHER INFORMATION: compounds."
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 942..7115
/ OTHER INFORMATION: /product= "Module 1 of SorA"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 7203..12884
/ OTHER INFORMATION: /product= "Module 2 of SorA"

```

FEATURE:
NAME/KEY: misc feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of SorA"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "SorB"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs genes"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "SorM"
OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin."
US-08-764-233A-1
Query Match 63.2%; Score 15.8; DB 1; Length 49377;
Best Local Similarity 89.5%; Pred. No. 5e+02; Mismatches 0; Gaps 0;
Matches 17; Conservative 0;
QY 1 CCCGAGGGCCAGAGGGCA 19
DB 15287 CCTGAGGCCAGAGGGCA 15269
RESULT 25
US-09-621-976-1421/c
Sequence 1421, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621.976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19355
SOFTWARE: Patent.pm
SEQ ID NO 1421
LENGTH: 539
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 243..467
NAME/KEY: sig peptide
LOCATION: 243..437
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 8.19999980926514

OTHER INFORMATION: seq LEFLIQLVTVARG/CW
US-09-621-976-1421
Query Match 62.4%; Score 15.6; DB 4; Length 539;
Best Local Similarity 81.8%; Pred. No. 5.4e+02; Mismatches 18; Conservative 0; Gaps 0;
QY 4 GGAGGCCAGAGGGCACATAGC 25
DB 381 GGAGCAACAGAGGGACATAGC 360
RESULT 26
US-09-719-919A-2/c
Sequence 2, Application US/09719919A
Patent No. 6680180
GENERAL INFORMATION:
APPLICANT: Jegla, Timothy James
APPLICANT: ICAGEN Incorporated
TITLE OF INVENTION: Kv6.2, a Voltage-Gated Potassium Channel Subunit
FILE REFERENCE: 018512-001410US
CURRENT APPLICATION NUMBER: US/09/719.919A
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: US 60/091,466
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: WO PCT/US99/14945
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1518
TYPE: DNA
ORGANISM: Mus sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1518)
OTHER INFORMATION: mouse alpha subunit of heteromeric voltage-gated potassium channel Kv6.2
US-09-719-919A-2
Query Match 62.4%; Score 15.6; DB 4; Length 1518;
Best Local Similarity 81.8%; Pred. No. 5.6e+02; Mismatches 18; Conservative 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCACAT 22
DB 430 CCCGAGGACAGAGGGCACAT 409
RESULT 27
US-09-061-154-1/c
Sequence 1, Application US/09061154
Patent No. 6462188
GENERAL INFORMATION:
APPLICANT: Kirkness, E.
TITLE OF INVENTION: NEW HUMAN 5-HT3 RECEPTOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/061,154
FILING DATE: 16-APR-1998
CLASSIFICATION: 536

```

;
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24366-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 67...1389
; OTHER INFORMATION:
;
; US-09-061-154-1
;
; Query Match 62.4%; Score 15.6; DB 4; Length 1613;
; Best Local Similarity 81.8%; Pred. No. 5.6e+02;
; Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 3 CGGAGGCCAGAGGCACATAG 24
; DB 415 CGGGGGCCAGATGGCACTTAG 394
;
; RESULT 29
; US-08-991-426-3
; Sequence 3, Application US/08991426
; Patent No. 6013257
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,426
; FILING DATE: 16-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/851,160
; FILING DATE: 05-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/643,798
; FILING DATE: 07-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/043001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 86...1276
;
; US-08-991-426-3
;
; Query Match 62.4%; Score 15.6; DB 3; Length 1654;
; Best Local Similarity 81.8%; Pred. No. 5.6e+02;
; Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 4 GGAGGGCCAGAGGCACATAGC 25
; DB 617 GGAGGGCCCTGTGGCAGCGAGC 638
;
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24366-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 67...1389
; OTHER INFORMATION:
;
; US-09-061-154-3/c
; Sequence 3, Application US/09061154
; Patent No. 6462188
; GENERAL INFORMATION:
; APPLICANT: Kirkness, E.
; TITLE OF INVENTION: NEW HUMAN 5-HT3 RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/061,154
; FILING DATE: 16-APR-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24366-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```



```
RESULT 30
US-09-143-470-3
; Sequence 3, Application US/09143470
; Patent No. 6043086
; GENERAL INFORMATION:
; APPLICANT: Pal, Yang
; TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR
; FILE REFERENCE: 09404/049001
; CURRENT APPLICATION NUMBER: US/09/143,470
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1654
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (86)...(1276)
US-09-143-470-3
Query Match      62.4%; Score 15.6; DB 3; Length 1654;
Best Local Similarity 81.8%; Pred. No. 5.6e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      4 GGAGGGCCAGAGGGCACATAGC 25
        |||||
Db      617 GGAGGGCCTGTGGCAGCGAGC 638

Search completed: June 20, 2004, 11:47:06
Job time : 39.7705 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 141.402 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccggagggcagggcacatagc 25

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/US06_PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 25 | 100.0 | 1584 | 9 | US-09-712-363-10 |
| 2 | 20.2 | 80.8 | 276820 | 13 | US-10-271-416-9 |
| 3 | 18.8 | 75.2 | 813 | 13 | US-10-027-632-144075 |
| 4 | 18.8 | 75.2 | 813 | 16 | US-10-027-632-144075 |
| 5 | 18.6 | 74.4 | 59817 | 12 | US-10-052-482-190 |
| 6 | 18.4 | 73.6 | 14645 | 10 | US-09-960-706-1043 |
| 7 | 18.4 | 73.6 | 14646 | 10 | US-09-873-319-691 |
| 8 | 18.2 | 72.8 | 90442 | 15 | US-10-105-637-1 |
| 9 | 18.2 | 72.8 | 90442 | 16 | US-10-034-650-43 |
| 10 | 17.6 | 70.4 | 523 | 15 | US-10-029-386-11131 |
| 11 | 17.6 | 70.4 | 637 | 10 | US-09-791-279-49 |
| 12 | 17.6 | 70.4 | 834 | 13 | US-10-027-632-140370 |
| 13 | 17.6 | 70.4 | 834 | 16 | US-10-027-632-140370 |
| 14 | 17.6 | 70.4 | 1010 | 13 | US-10-027-632-251702 |

| | | | | | | |
|----|------|------|--------|----|----------------------|---------------------|
| 15 | 17.6 | 70.4 | 1010 | 16 | US-10-027-632-351702 | Sequence 251702, Ap |
| 16 | 17.6 | 70.4 | 1152 | 15 | US-10-017-161-1679 | Sequence 1679, Ap |
| 17 | 17.6 | 70.4 | 1152 | 16 | US-10-292-798-1339 | Sequence 1339, Ap |
| 18 | 17.6 | 70.4 | 2545 | 13 | US-10-027-632-112152 | Sequence 112152, Ap |
| 19 | 17.6 | 70.4 | 2545 | 13 | US-10-027-632-112153 | Sequence 112153, Ap |
| 20 | 17.6 | 70.4 | 2545 | 16 | US-10-027-632-112152 | Sequence 112152, Ap |
| 21 | 17.6 | 70.4 | 2545 | 16 | US-10-027-632-112153 | Sequence 112153, Ap |
| 22 | 17.4 | 69.6 | 24902 | 13 | US-10-087-192-1045 | Sequence 1045, Ap |
| 23 | 17.2 | 68.8 | 440 | 13 | US-10-424-599-139706 | Sequence 139706, Ap |
| 24 | 17.2 | 68.8 | 733 | 13 | US-10-027-632-144076 | Sequence 144076, Ap |
| 25 | 17.2 | 68.8 | 733 | 16 | US-10-027-632-144076 | Sequence 144076, Ap |
| 26 | 17.2 | 68.8 | 3014 | 16 | US-10-429-223-1 | Sequence 1, Appli |
| 27 | 17.2 | 68.8 | 23070 | 12 | US-10-052-482-13 | Sequence 13, Appli |
| 28 | 17 | 68.0 | 376 | 9 | US-09-925-302-409 | Sequence 409, App |
| 29 | 17 | 68.0 | 376 | 13 | US-09-925-302-409 | Sequence 409, App |
| 30 | 17 | 68.0 | 406 | 10 | US-09-918-995-6988 | Sequence 6988, Ap |
| 31 | 17 | 68.0 | 437 | 9 | US-09-925-302-402 | Sequence 402, App |
| 32 | 17 | 68.0 | 437 | 13 | US-09-925-302-402 | Sequence 402, App |
| 33 | 17 | 68.0 | 547 | 17 | US-10-343-710-49 | Sequence 49, Appli |
| 34 | 17 | 68.0 | 552 | 9 | US-09-864-761-9652 | Sequence 9652, Ap |
| 35 | 17 | 68.0 | 655 | 13 | US-10-027-632-103742 | Sequence 103742, Ap |
| 36 | 17 | 68.0 | 655 | 16 | US-10-027-632-103742 | Sequence 103742, Ap |
| 37 | 17 | 68.0 | 681 | 12 | US-10-336-603A-97 | Sequence 97, Appli |
| 38 | 17 | 68.0 | 935 | 10 | US-09-971-429B-3 | Sequence 3, Appli |
| 39 | 17 | 68.0 | 997 | 9 | US-09-864-761-9278 | Sequence 9278, Ap |
| 40 | 17 | 68.0 | 1161 | 12 | US-10-336-603A-99 | Sequence 99, Appli |
| 41 | 17 | 68.0 | 1836 | 13 | US-10-282-122A-27839 | Sequence 27839, A |
| 42 | 17 | 68.0 | 2248 | 9 | US-09-925-302-30 | Sequence 30, Appli |
| 43 | 17 | 68.0 | 2248 | 13 | US-09-925-302-30 | Sequence 30, Appli |
| 44 | 17 | 68.0 | 2487 | 13 | US-10-027-632-103365 | Sequence 103365, Ap |
| 45 | 17 | 68.0 | 2487 | 16 | US-10-027-632-103365 | Sequence 103365, Ap |
| 46 | 17 | 68.0 | 4732 | 9 | US-09-954-458-2079 | Sequence 2079, Ap |
| 47 | 17 | 68.0 | 4732 | 10 | US-09-960-706-699 | Sequence 699, App |
| 48 | 17 | 68.0 | 4732 | 10 | US-09-873-319-441 | Sequence 441, App |
| 49 | 17 | 68.0 | 4778 | 9 | US-09-954-458-2031 | Sequence 2031, Ap |
| 50 | 17 | 68.0 | 5438 | 17 | US-10-601-807-3 | Sequence 3, Appli |
| 51 | 17 | 68.0 | 21732 | 9 | US-09-764-847-1414 | Sequence 1414, Ap |
| 52 | 17 | 68.0 | 21732 | 15 | US-10-092-154-1414 | Sequence 1414, Ap |
| 53 | 17 | 68.0 | 131576 | 13 | US-10-087-152-1564 | Sequence 1564, Ap |
| 54 | 17 | 68.0 | 174424 | 9 | US-09-967-768A-314 | Sequence 314, App |
| 55 | 17 | 68.0 | 174424 | 10 | US-09-960-706-969 | Sequence 969, App |
| 56 | 17 | 68.0 | 186510 | 15 | US-10-043-715-1 | Sequence 1, Appli |
| 57 | 17 | 68.0 | 201143 | 13 | US-10-240-425-1099 | Sequence 1099, Ap |
| 58 | 17 | 68.0 | 493631 | 13 | US-10-087-192-205 | Sequence 205, App |
| 59 | 16.8 | 67.2 | 462 | 11 | US-09-801-944B-22 | Sequence 22, Appli |
| 60 | 16.8 | 67.2 | 464 | 10 | US-09-918-995-31310 | Sequence 31310, A |
| 61 | 16.8 | 67.2 | 480 | 13 | US-10-027-632-276278 | Sequence 276278, Ap |
| 62 | 16.8 | 67.2 | 480 | 16 | US-10-027-632-276278 | Sequence 276278, Ap |
| 63 | 16.8 | 67.2 | 551 | 13 | US-10-027-632-13200 | Sequence 13200, A |
| 64 | 16.8 | 67.2 | 551 | 13 | US-10-027-632-144828 | Sequence 144828, Ap |
| 65 | 16.8 | 67.2 | 551 | 16 | US-10-027-632-13200 | Sequence 13200, A |
| 66 | 16.8 | 67.2 | 551 | 16 | US-10-027-632-144828 | Sequence 144828, Ap |
| 67 | 16.8 | 67.2 | 729 | 13 | US-10-027-632-15719 | Sequence 15719, A |
| 68 | 16.8 | 67.2 | 729 | 16 | US-10-027-632-15719 | Sequence 15719, A |
| 69 | 16.8 | 67.2 | 879 | 15 | US-10-198-846-1415 | Sequence 1415, Ap |
| 70 | 16.8 | 67.2 | 921 | 15 | US-10-198-846-10634 | Sequence 10634, A |
| 71 | 16.8 | 67.2 | 988 | 13 | US-10-424-599-70641 | Sequence 70641, A |
| 72 | 16.8 | 67.2 | 1317 | 9 | US-09-822-830A-180 | Sequence 180, App |
| 73 | 16.8 | 67.2 | 1326 | 13 | US-10-342-897-1772 | Sequence 1772, Ap |
| 74 | 16.8 | 67.2 | 1326 | 13 | US-10-176-118-1772 | Sequence 1772, Ap |
| 75 | 16.8 | 67.2 | 1976 | 13 | US-10-276-774-817 | Sequence 817, App |
| 76 | 16.8 | 67.2 | 1979 | 13 | US-10-116-802-227 | Sequence 227, App |
| 77 | 16.8 | 67.2 | 2017 | 15 | US-10-198-846-9685 | Sequence 9685, Ap |
| 78 | 16.8 | 67.2 | 2038 | 10 | US-09-971-429B-35 | Sequence 35, Appli |
| 79 | 16.8 | 67.2 | 2038 | 14 | US-10-002-600-29 | Sequence 29, Appli |
| 80 | 16.8 | 67.2 | 2038 | 15 | US-10-101-510-606 | Sequence 606, App |
| 81 | 16.8 | 67.2 | 2059 | 15 | US-10-091-333-5 | Sequence 5, Appli |
| 82 | 16.8 | 67.2 | 2059 | 15 | US-10-325-878-5 | Sequence 5, Appli |
| 83 | 16.8 | 67.2 | 2942 | 16 | US-10-108-260A-599 | Sequence 599, App |
| 84 | 16.8 | 67.2 | 3156 | 13 | US-10-302-172-870 | Sequence 870, App |
| 85 | 16.8 | 67.2 | 3458 | 16 | US-10-094-749-927 | Sequence 927, App |
| 86 | 16.8 | 67.2 | 6584 | 13 | US-09-973-278-710 | Sequence 710, App |
| 87 | 16.8 | 67.2 | 13240 | 15 | US-10-017-161-1673 | Sequence 1673, Ap |

| | | | | | | | | | | | | | |
|-----|------|------|---------|----|----------------------|--------------------|-------|----|------|------|----|----------------------|--------------------|
| 88 | 16.8 | 67.2 | 13240 | 16 | US-10-292-798-1333 | Sequence 1333, Ap | C 161 | 16 | 64.0 | 323 | 16 | US-10-242-535A-6708 | Sequence 6708, Ap |
| 89 | 16.8 | 67.2 | 13830 | 15 | US-10-017-161-1581 | Sequence 1581, Ap | C 162 | 16 | 64.0 | 431 | 9 | US-09-983-965-4592 | Sequence 4592, Ap |
| 90 | 16.8 | 67.2 | 13830 | 16 | US-10-292-798-1257 | Sequence 1257, Ap | C 163 | 16 | 64.0 | 438 | 9 | US-09-864-711-5 | Sequence 5, Appl |
| 91 | 16.8 | 67.2 | 12524 | 16 | US-10-292-798-1255 | Sequence 1255, Ap | C 164 | 16 | 64.0 | 444 | 9 | US-09-864-761-3157 | Sequence 3157, Ap |
| 92 | 16.8 | 67.2 | 87415 | 13 | US-10-087-192-1651 | Sequence 1651, Ap | C 165 | 16 | 64.0 | 459 | 13 | US-10-085-783A-43067 | Sequence 43067, A |
| 93 | 16.6 | 66.4 | 364 | 9 | US-09-925-297-376 | Sequence 376, App | C 166 | 16 | 64.0 | 459 | 16 | US-10-242-535A-43067 | Sequence 43067, A |
| 94 | 16.6 | 66.4 | 700 | 13 | US-10-027-632-162875 | Sequence 162875, A | C 167 | 16 | 64.0 | 460 | 10 | US-09-918-995-28156 | Sequence 28156, A |
| 95 | 16.6 | 66.4 | 700 | 16 | US-10-027-632-162875 | Sequence 162875, A | C 168 | 16 | 64.0 | 460 | 10 | US-10-060-036-3048 | Sequence 3048, Ap |
| 96 | 16.6 | 66.4 | 725 | 13 | US-10-027-632-103842 | Sequence 103842, A | C 169 | 16 | 64.0 | 473 | 13 | US-10-085-783A-35519 | Sequence 35519, A |
| 97 | 16.6 | 66.4 | 725 | 13 | US-10-027-632-323554 | Sequence 323554, A | C 170 | 16 | 64.0 | 473 | 13 | US-10-242-535A-35519 | Sequence 35519, A |
| 98 | 16.6 | 66.4 | 725 | 16 | US-10-027-632-103842 | Sequence 103842, A | C 171 | 16 | 64.0 | 490 | 9 | US-09-925-297-172 | Sequence 172, App |
| 99 | 16.6 | 66.4 | 725 | 16 | US-10-027-632-323554 | Sequence 323554, A | C 172 | 16 | 64.0 | 490 | 15 | US-10-060-036-3697 | Sequence 3697, Ap |
| 100 | 16.6 | 66.4 | 761 | 13 | US-10-027-632-30195 | Sequence 30195, A | C 173 | 16 | 64.0 | 516 | 10 | US-09-770-961-116 | Sequence 116, App |
| 101 | 16.6 | 66.4 | 761 | 16 | US-10-027-632-30195 | Sequence 30195, A | C 174 | 16 | 64.0 | 525 | 15 | US-10-029-386-20660 | Sequence 20660, A |
| 102 | 16.6 | 66.4 | 765 | 13 | US-10-027-632-103927 | Sequence 103927, A | C 175 | 16 | 64.0 | 528 | 12 | US-09-963-034-917 | Sequence 917, App |
| 103 | 16.6 | 66.4 | 765 | 13 | US-10-027-632-104151 | Sequence 104151, A | C 176 | 16 | 64.0 | 541 | 15 | US-10-023-386-3050 | Sequence 9050, Ap |
| 104 | 16.6 | 66.4 | 765 | 13 | US-10-027-632-323670 | Sequence 323670, A | C 177 | 16 | 64.0 | 555 | 13 | US-10-029-386-22402 | Sequence 22402, A |
| 105 | 16.6 | 66.4 | 765 | 13 | US-10-027-632-325048 | Sequence 325048, A | C 178 | 16 | 64.0 | 555 | 15 | US-10-023-386-24654 | Sequence 24654, A |
| 106 | 16.6 | 66.4 | 765 | 16 | US-10-027-632-103927 | Sequence 103927, A | C 179 | 16 | 64.0 | 567 | 13 | US-10-029-386-24654 | Sequence 24654, A |
| 107 | 16.6 | 66.4 | 765 | 16 | US-10-027-632-104151 | Sequence 104151, A | C 180 | 16 | 64.0 | 567 | 13 | US-10-027-632-274762 | Sequence 274762, A |
| 108 | 16.6 | 66.4 | 765 | 16 | US-10-027-632-323670 | Sequence 323670, A | C 181 | 16 | 64.0 | 567 | 16 | US-10-027-632-274762 | Sequence 274762, A |
| 109 | 16.6 | 66.4 | 835 | 13 | US-10-027-632-325048 | Sequence 325048, A | C 182 | 16 | 64.0 | 567 | 16 | US-10-027-632-274762 | Sequence 274762, A |
| 110 | 16.6 | 66.4 | 835 | 16 | US-10-027-632-152146 | Sequence 152146, A | C 183 | 16 | 64.0 | 584 | 15 | US-10-027-632-274762 | Sequence 274762, A |
| 111 | 16.6 | 66.4 | 835 | 16 | US-10-027-632-152146 | Sequence 152146, A | C 184 | 16 | 64.0 | 584 | 15 | US-10-001-883-49 | Sequence 49, Appl |
| 112 | 16.6 | 66.4 | 1129 | 16 | US-10-369-493-26822 | Sequence 26822, A | C 185 | 16 | 64.0 | 584 | 9 | US-09-864-761-8182 | Sequence 8182, Ap |
| 113 | 16.6 | 66.4 | 1774 | 15 | US-10-261-078-20 | Sequence 20, Appl | C 186 | 16 | 64.0 | 590 | 15 | US-09-864-761-13596 | Sequence 13596, A |
| 114 | 16.6 | 66.4 | 11987 | 10 | US-08-764-891-9972 | Sequence 9972, Ap | C 187 | 16 | 64.0 | 590 | 15 | US-10-029-386-9674 | Sequence 8674, Ap |
| 115 | 16.6 | 66.4 | 12575 | 10 | US-08-764-891-9971 | Sequence 9971, Ap | C 188 | 16 | 64.0 | 596 | 15 | US-10-029-386-10951 | Sequence 10951, A |
| 116 | 16.6 | 66.4 | 43793 | 12 | US-09-997-722-37 | Sequence 37, Appl | C 189 | 16 | 64.0 | 607 | 9 | US-09-809-545A-19 | Sequence 19, Appl |
| 117 | 16.6 | 66.4 | 256493 | 13 | US-10-087-192-1000 | Sequence 1000, Ap | C 190 | 16 | 64.0 | 608 | 13 | US-10-027-632-71049 | Sequence 71049, A |
| 118 | 16.6 | 66.4 | 3186778 | 13 | US-10-027-632-174961 | Sequence 174961, A | C 191 | 16 | 64.0 | 608 | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 119 | 16.6 | 66.4 | 3186778 | 16 | US-10-027-632-174961 | Sequence 174961, A | C 192 | 16 | 64.0 | 608 | 16 | US-10-027-632-71049 | Sequence 71049, A |
| 120 | 16.4 | 65.6 | 261 | 11 | US-09-864-408A-2621 | Sequence 2621, Ap | C 193 | 16 | 64.0 | 609 | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 121 | 16.4 | 65.6 | 1494 | 14 | US-10-119-635-3 | Sequence 3, Appl | C 194 | 16 | 64.0 | 609 | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 122 | 16.4 | 65.6 | 1350 | 14 | US-10-016-157A-126 | Sequence 126, App | C 195 | 16 | 64.0 | 609 | 16 | US-10-027-632-312331 | Sequence 312331, A |
| 123 | 16.4 | 65.6 | 2240 | 13 | US-10-119-635-7 | Sequence 7, Appl | C 196 | 16 | 64.0 | 617 | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 124 | 16.4 | 65.6 | 2240 | 14 | US-10-213-796-19 | Sequence 19, Appl | C 197 | 16 | 64.0 | 617 | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 125 | 16.4 | 65.6 | 51664 | 13 | US-10-087-192-877 | Sequence 877, Ap | C 198 | 16 | 64.0 | 617 | 16 | US-10-027-632-312331 | Sequence 312331, A |
| 126 | 16.4 | 65.6 | 94672 | 13 | US-10-087-192-1069 | Sequence 1069, Ap | C 199 | 16 | 64.0 | 617 | 16 | US-10-027-632-312331 | Sequence 312331, A |
| 127 | 16.4 | 65.6 | 106315 | 13 | US-10-087-192-178 | Sequence 178, Ap | C 200 | 16 | 64.0 | 621 | 13 | US-10-282-122A-26617 | Sequence 26617, A |
| 128 | 16.2 | 64.8 | 186 | 15 | US-10-029-386-14920 | Sequence 14920, A | C 201 | 16 | 64.0 | 624 | 16 | US-10-369-493-31482 | Sequence 31482, A |
| 129 | 16.2 | 64.8 | 433 | 13 | US-10-027-632-181988 | Sequence 181988, A | C 202 | 16 | 64.0 | 625 | 13 | US-10-027-632-223115 | Sequence 223115, A |
| 130 | 16.2 | 64.8 | 433 | 16 | US-10-027-632-181988 | Sequence 181988, A | C 203 | 16 | 64.0 | 625 | 16 | US-10-027-632-223115 | Sequence 223115, A |
| 131 | 16.2 | 64.8 | 473 | 12 | US-09-876-143-427 | Sequence 427, App | C 204 | 16 | 64.0 | 634 | 13 | US-10-027-632-134639 | Sequence 134639, A |
| 132 | 16.2 | 64.8 | 542 | 13 | US-10-029-386-1218 | Sequence 1218, Ap | C 205 | 16 | 64.0 | 634 | 16 | US-10-027-632-134639 | Sequence 134639, A |
| 133 | 16.2 | 64.8 | 543 | 13 | US-10-424-599-83481 | Sequence 83481, A | C 206 | 16 | 64.0 | 662 | 13 | US-10-027-632-48022 | Sequence 48022, A |
| 134 | 16.2 | 64.8 | 892 | 13 | US-10-027-632-155735 | Sequence 155735, A | C 207 | 16 | 64.0 | 662 | 16 | US-10-027-632-48022 | Sequence 48022, A |
| 135 | 16.2 | 64.8 | 892 | 16 | US-10-027-632-155735 | Sequence 155735, A | C 208 | 16 | 64.0 | 689 | 10 | US-09-791-279-20 | Sequence 20, Appl |
| 136 | 16.2 | 64.8 | 898 | 13 | US-10-027-632-120372 | Sequence 120372, A | C 209 | 16 | 64.0 | 691 | 15 | US-10-029-386-32763 | Sequence 32763, A |
| 137 | 16.2 | 64.8 | 898 | 16 | US-10-027-632-120372 | Sequence 120372, A | C 210 | 16 | 64.0 | 730 | 13 | US-10-029-386-32763 | Sequence 32763, A |
| 138 | 16.2 | 64.8 | 931 | 16 | US-10-050-704-88 | Sequence 88, Appl | C 211 | 16 | 64.0 | 730 | 16 | US-10-027-632-137926 | Sequence 137926, A |
| 139 | 16.2 | 64.8 | 940 | 15 | US-10-050-704-48 | Sequence 48, Appl | C 212 | 16 | 64.0 | 733 | 13 | US-10-027-632-17850 | Sequence 17850, A |
| 140 | 16.2 | 64.8 | 1262 | 10 | US-09-984-271-44 | Sequence 44, Appl | C 213 | 16 | 64.0 | 733 | 16 | US-10-027-632-17850 | Sequence 17850, A |
| 141 | 16.2 | 64.8 | 1262 | 13 | US-09-984-276-44 | Sequence 44, Appl | C 214 | 16 | 64.0 | 747 | 13 | US-10-027-632-36534 | Sequence 26534, A |
| 142 | 16.2 | 64.8 | 2945 | 15 | US-10-106-598-228 | Sequence 228, App | C 215 | 16 | 64.0 | 747 | 16 | US-10-027-632-36534 | Sequence 26534, A |
| 143 | 16.2 | 64.8 | 471 | 10 | US-09-764-891-9043 | Sequence 9043, Ap | C 216 | 16 | 64.0 | 925 | 15 | US-10-017-161-1503 | Sequence 1503, Ap |
| 144 | 16.2 | 64.8 | 5033 | 16 | US-10-369-300-14 | Sequence 14, Appl | C 217 | 16 | 64.0 | 1149 | 13 | US-10-424-589-16392 | Sequence 16392, A |
| 145 | 16.2 | 64.8 | 5610 | 15 | US-10-157-031-127 | Sequence 127, App | C 218 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7779 | Sequence 7779, Ap |
| 146 | 16.2 | 64.8 | 47309 | 16 | US-10-085-117-79 | Sequence 79, Appl | C 219 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7781 | Sequence 7781, Ap |
| 147 | 16.2 | 64.8 | 66972 | 13 | US-10-087-192-556 | Sequence 556, App | C 220 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7905 | Sequence 7905, Ap |
| 148 | 16.2 | 64.8 | 146547 | 15 | US-10-017-128-1 | Sequence 1, Appl | C 221 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7907 | Sequence 7907, Ap |
| 149 | 16.2 | 64.8 | 168821 | 13 | US-10-087-192-622 | Sequence 622, App | C 222 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7909 | Sequence 7909, Ap |
| 150 | 16.2 | 64.8 | 348548 | 13 | US-10-087-192-334 | Sequence 334, App | C 223 | 16 | 64.0 | 1197 | 15 | US-10-087-268-3 | Sequence 3, Appl |
| 151 | 16 | 64.0 | 17 | 17 | US-10-712-672-2616 | Sequence 2616, Ap | C 224 | 16 | 64.0 | 1278 | 15 | US-10-087-268-6 | Sequence 6, Appl |
| 152 | 16 | 64.0 | 172 | 13 | US-10-085-783A-5188 | Sequence 5188, Ap | C 225 | 16 | 64.0 | 1348 | 17 | US-10-343-710-4 | Sequence 14, Appl |
| 153 | 16 | 64.0 | 172 | 16 | US-10-242-535A-5188 | Sequence 5188, Ap | C 226 | 16 | 64.0 | 1400 | 16 | US-10-330-051A-36 | Sequence 36, Appl |
| 154 | 16 | 64.0 | 201 | 9 | US-09-983-965-5267 | Sequence 5267, A | C 227 | 16 | 64.0 | 1400 | 17 | US-10-343-710-35 | Sequence 35, Appl |
| 155 | 16 | 64.0 | 230 | 13 | US-10-085-783A-13031 | Sequence 13031, A | C 228 | 16 | 64.0 | 1497 | 17 | US-10-343-710-35 | Sequence 35, Appl |
| 156 | 16 | 64.0 | 230 | 16 | US-10-242-535A-13031 | Sequence 13031, A | C 229 | 16 | 64.0 | 1544 | 9 | US-09-349-015-18 | Sequence 18, Appl |
| 157 | 16 | 64.0 | 294 | 9 | US-09-864-761-19936 | Sequence 19936, A | C 230 | 16 | 64.0 | 1544 | 15 | US-10-219-664-14 | Sequence 14, Appl |
| 158 | 16 | 64.0 | 298 | 16 | US-10-330-051A-30 | Sequence 30, Appl | C 231 | 16 | 64.0 | 1583 | 16 | US-10-292-798-1205 | Sequence 1205, Ap |
| 159 | 16 | 64.0 | 321 | 13 | US-10-424-599-27398 | Sequence 27398, A | C 232 | 16 | 64.0 | 1604 | 13 | US-10-027-632-100159 | Sequence 100159, A |
| 160 | 16 | 64.0 | 323 | 13 | US-10-085-783A-6708 | Sequence 6708, Ap | C 233 | 16 | 64.0 | 1604 | 16 | US-10-027-632-100159 | Sequence 100159, A |

Sequence 12610, A
Sequence 448, App
Sequence 15, Appl
Sequence 659, App
Sequence 97355, A
Sequence 97355, A
Sequence 11, Appl
Sequence 10088, A
Sequence 3291, App
Sequence 872, App
Sequence 872, App
Sequence 3, Appl
Sequence 1, Appl
Sequence 255, App
Sequence 961, App
Sequence 112241, App
Sequence 112241, App
Sequence 71, Appl
Sequence 1683, App
Sequence 43549, A
Sequence 115939, App
Sequence 115940, App
Sequence 115941, App
Sequence 115939, App
Sequence 115940, App
Sequence 115941, App
Sequence 341, App
Sequence 106, App
Sequence 2353, App
Sequence 35, Appl
Sequence 40, Appl
Sequence 3, Appl
Sequence 5491, App
Sequence 1, Appl
Sequence 775, App
Sequence 106, App
Sequence 3, Appl
Sequence 3, Appl
Sequence 217, App
Sequence 1102, App
Sequence 22, Appl
Sequence 241, App
Sequence 853, App
Sequence 1282, App
Sequence 1990, App
Sequence 423, App
Sequence 81174, A
Sequence 58, Appl
Sequence 3239, App
Sequence 3239, App
Sequence 16851, A
Sequence 16851, A
Sequence 289524, App
Sequence 289524, App
Sequence 17, Appl
Sequence 285723, App
Sequence 285723, App
Sequence 13129, A
Sequence 103268, App
Sequence 103268, App
Sequence 16502, A
Sequence 16502, A
Sequence 132497, App
Sequence 132497, App
Sequence 13459, A
Sequence 13459, A

US-10-282-122A-12610
US-10-296-115-448
US-10-343-710-15
US-10-094-749-659
US-10-027-632-97355
US-10-027-632-97355
US-10-468-334-11
US-10-198-846-10088
US-09-880-107-3291
US-09-938-842A-872
US-09-938-842A-872
US-09-799-514-3
US-10-173-519-1
US-09-964-824A-255
US-09-954-531-961
US-10-027-632-112241
US-10-027-632-112241
US-10-403-161-71
US-10-017-161-1683
US-10-369-493-43549
US-10-027-632-115939
US-10-027-632-115940
US-10-027-632-115941
US-10-044-090-341
US-10-320-786-106
US-09-880-107-2353
US-10-372-683-35
US-10-333-314-40
US-10-601-807-3
US-09-764-891-5491
US-10-016-248-1
US-10-087-192-775
US-10-087-192-106
US-09-729-094-3
US-10-435-631-3
US-10-087-192-217
US-10-085-117-76
US-10-240-425-1102
US-10-087-192-22
US-10-085-117-241
US-10-087-192-853
US-10-087-192-1282
US-10-087-192-1990
US-09-933-797-423
US-10-424-599-81174
US-10-276-774-58
US-09-764-877-3239
US-10-242-515-3239
US-10-085-783A-16851
US-10-242-535A-16851
US-10-027-632-289524
US-10-027-632-289524
US-10-060-841-17
US-10-027-632-285723
US-10-027-632-285723
US-10-029-386-13129
US-10-027-632-103268
US-10-027-632-103268
US-10-027-632-16502
US-10-027-632-16502
US-10-027-632-132497
US-10-027-632-132497
US-10-027-632-13459
US-10-027-632-13459

ALIGNMENTS

US-09-712-363-10/c
; Sequence 10, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-10

Query Match 100.0%; Score 25; DB 9; Length 1584;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGGGCCACATAGC 25
DB 1019 CCGGAGGGCCAGGGCCACATAGC 995

RESULT 2
US-10-271-416-9/c
; Sequence 9, Application US/10271416
; Publication No. US20040043021A1
; GENERAL INFORMATION:
; APPLICANT: Keith, Tim
; APPLICANT: Little, Randall D.
; APPLICANT: Van Erdewegh, Paul
; APPLICANT: Dupuis, Josee
; APPLICANT: Del Mastro, Richard G.
; APPLICANT: Allen, Kristina
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4045
; CURRENT APPLICATION NUMBER: US/10/271,416
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: 60/328,424
; PRIOR FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 276820
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: conflict

; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1043
; LENGTH: 14646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X03100
; NAME/KEY: unsure
; LOCATION: (1)..(14646)
; OTHER INFORMATION: n = a or c or g or t
US-09-960-706-1043

Query Match 73.6%; Score 18.4; DB 10; Length 14646;
Best Local Similarity 95.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GAGGCCAGAGGGCACATAG 24
Db 2885 GAGGCCAGAGGGACATAG 2866

RESULT 7
US-09-873-319-691/c
; Sequence 691, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 691
; LENGTH: 14646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X03100
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(14646)
; OTHER INFORMATION: n = a or c or g or t
US-09-873-319-691

Query Match 73.6%; Score 18.4; DB 10; Length 14646;
Best Local Similarity 95.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GAGGCCAGAGGGCACATAG 24
Db 2885 GAGGCCAGAGGGACATAG 2866

RESULT 8
US-10-105-637-1
; Sequence 1, Application US/10105637
; Publication No. US2003008752A1
; GENERAL INFORMATION:

; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; TITLE OF INVENTION: ASSOCIATED WITH ALTERED EXPRESSION OF PRDM11
; FILE REFERENCE: 529452500120
; CURRENT APPLICATION NUMBER: US/10/105,637
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 10/034,650
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 90442
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-105-637-1

Query Match 72.8%; Score 18.2; DB 15; Length 90442;
Best Local Similarity 87.0%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCGAGGGCCAGAGGGCACATAG 24
Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 9
US-10-034-650-43
; Sequence 43, Application US/10034650
; Publication No. US20030216558A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000128
; CURRENT APPLICATION NUMBER: US/10/034,650
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 09/474,377
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 90442
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-034-650-43

Query Match 72.8%; Score 18.2; DB 16; Length 90442;
Best Local Similarity 87.0%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCGAGGGCCAGAGGGCACATAG 24
Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 10
US-10-029-386-11131/c
; Sequence 11131, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11131
LENGTH: 523
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004178.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.51
OTHER INFORMATION: SWISSPROT HIT: P15410, EVALUE 5.00e-23
OTHER INFORMATION: EST HUMAN HIT: BF883760.1, EVALUE 7.30e-01
OTHER INFORMATION: NT HIT: S78653.1, EVALUE 3.20e-02
US-10-029-386-11131

Query Match 70.4%; Score 17.6; DB 15; Length 523;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCAGAGGGGCACATAG 24
DB 413 CCAGTGGGACAGAGGGGCACAG 390

RESULT 11
US-09-791-279-49
Sequence 49, Application US/09791279
Publication No. US20030050456A1
GENERAL INFORMATION:
APPLICANT: Vogeli, Gabriel
APPLICANT: Wood, Linda S.
APPLICANT: Parodi Luis
APPLICANT: Lind, Peter
TITLE OF INVENTION: No. US20030050456A1 G Protein-Coupled Receptors
FILE REFERENCE: 00048.US1
CURRENT APPLICATION NUMBER: US/09/791,279
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: 60/184,715
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,725
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,712
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,606
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,602
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,604
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,822
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,710
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,689
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,690
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,716
PRIOR FILING DATE: 2000-02-24
NUMBER OF SEQ ID NOS: 220
SOFTWARE: PatentIn version 3.0
SEQ ID NO 49
LENGTH: 637
TYPE: DNA
ORGANISM: Homo sapiens
US-09-791-279-49

Query Match 70.4%; Score 17.6; DB 10; Length 637;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCAGAGGGGCACATAG 24
DB 182 CCAGTGGGACAGAGGGGCACAG 205

RESULT 12
US-10-027-632-140370/c
Sequence 140370, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 140370
LENGTH: 834
TYPE: DNA
ORGANISM: Human
US-10-027-632-140370

Query Match 70.4%; Score 17.6; DB 13; Length 834;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCAGAGGGGCACATAG 24
DB 154 CCAGAGGGACAGAGGGGCATGG 131

RESULT 13
US-10-027-632-140370/c
Sequence 140370, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720


```
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 140370
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140370

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 16; Length 834;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 154 CCCAGAGGACAGAGGGCAGATGG 131

RESULT 14
US-10-027-632-251702
; Sequence 251702, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 251702
; LENGTH: 1010
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251702

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 16; Length 1010;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 380 CCCGAGAGGCGAGAGGGCACACAG 403

RESULT 15
US-10-027-632-251702
; Sequence 251702, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 251702
; LENGTH: 1010
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251702

Query Match
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 380 CCCGAGAGGCGAGAGGGCACACAG 403

RESULT 16
US-10-017-161-1679/C
; Sequence 1679, Application US/10017161
; Publication No. US20030143688A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABEURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017.161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1679
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1152)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(592)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (781)..(952)
US-10-017-161-1679

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 15; Length 1152;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 554 CCAGGTGGGACAGAGGGCACAGAG 531

RESULT 17
US-10-292-798-1339/C
; Sequence 1339, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
```

APPLICANT: SUWA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 1339
LENGTH: 1152
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: source
FEATURE:
LOCATION: (1)..(1152)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(629)
FEATURE:
NAME/KEY: CDS
LOCATION: (728)..(952)
US-10-292-798-1339

Query Match 70.4%; Score 17.6; DB 16; Length 1152;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGCGACATAG 24
Db 554 CCAGGTGGACAGAGGCGACAGAG 531

RESULT 18
US-10-632-112152
Sequence 112152, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 112152
LENGTH: 2545
TYPE: DNA
ORGANISM: Human
US-10-632-112152

Query Match 70.4%; Score 17.6; DB 13; Length 2545;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGCGACATAG 24
Db 907 CCCAGAGGGACAGAGGCGACATGG 930

RESULT 19
US-10-632-112153
Sequence 112153, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 112153
LENGTH: 2545
TYPE: DNA
ORGANISM: Human
US-10-632-112153

Query Match 70.4%; Score 17.6; DB 13; Length 2545;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGCGACATAG 24
Db 907 CCCAGAGGGACAGAGGCGACATGG 930

RESULT 20
US-10-632-112152
Sequence 112152, Application US/10027632
Publication No. US20020204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112152
; LENGTH: 2545
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-112152

Query Match 70.4%; Score 17.6; DB 16; Length 2545;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCCACATAG 24
Db 907 CCAGAGGGACAGAGGGCCACATGG 930

RESULT 21
US-10-027-632-112153
; Sequence 112153, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112153
; LENGTH: 2545
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-112153

Query Match 70.4%; Score 17.6; DB 16; Length 2545;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCCACATAG 24
Db 907 CCAGAGGGACAGAGGGCCACATGG 930

RESULT 22
US-10-087-192-1045
; Sequence 1045, Application US/10087192
; Publication No. US2002018586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087.192
; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1045
; LENGTH: 24902
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...((24902))
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1045

Query Match 69.6%; Score 17.4; DB 13; Length 24902;
Best Local Similarity 94.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 AGGGCCAGAGGGCCACATAG 24
Db 504 AGGGCCAGAGGGCCACAGAG 522

RESULT 23
US-10-424-599-139706
; Sequence 139706, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 139706
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_97162C.1
US-10-424-599-139706

Query Match 68.8%; Score 17.2; DB 13; Length 440;
Best Local Similarity 86.4%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CCGAGGGCCAGAGGGCCACATAG 24
Db 109 CCGAGGGCCCTGTGAGCAGCATAG 130

RESULT 24
US-10-027-632-144076/c
; Sequence 144076, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144076
; LENGTH: 733
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144076

Query Match 68.8%; Score 17.2; DB 13; Length 733;
Best Local Similarity 86.4%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCACATA 23
DB 515 CAGGAAGGCCAGAGGGCACAAA 494

RESULT 25
US-10-027-632-144076/c
; Sequence 144076, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144076
; LENGTH: 733
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144076

Query Match 68.8%; Score 17.2; DB 16; Length 733;
Best Local Similarity 86.4%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCACATA 23
DB 515 CAGGAAGGCCAGAGGGCACAAA 494

RESULT 26
US-10-429-223-1
; Sequence 1, Application US/10429223
; Publication No. US20040009513A1
; GENERAL INFORMATION:

; APPLICANT: Vatner, Stephen
; APPLICANT: Depre, Christophe
; TITLE OF INVENTION: pDJAL, a cardiac specific gene,
; TITLE OF INVENTION: corresponding proteins and uses thereof
; FILE REFERENCE: 601-1-130
; CURRENT APPLICATION NUMBER: US/10/429,223
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: 60/377,578
; PRIOR FILING DATE: 2002-05-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3014
; TYPE: DNA
; ORGANISM: Sus scrofa
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; OTHER INFORMATION: n = A,T,C or G
US-10-429-223-1

Query Match 68.8%; Score 17.2; DB 16; Length 3014;
Best Local Similarity 86.4%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGAGGGCCAGAGGGCACAT 22
DB 2592 CCGAGGGCCAGAGGGCACAT 2613

RESULT 27
US-10-052-482-13
; Sequence 13, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 23070
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (370)..(389)
; OTHER INFORMATION: "n" at positions 370 to 389 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3127)..(3343)
; OTHER INFORMATION: "n" at positions 3127 to 3343 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (17723)..(17758)
; OTHER INFORMATION: "n" at positions 17723 to 17758 can be any base.
US-10-052-482-13

Query Match 68.8%; Score 17.2; DB 12; Length 23070;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCACATA 23
DB 18715 CCGAGGGCCAGAGGGCACATA 18736

RESULT 28

US-09-925-302-409/c
; Sequence 409, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 409
; LENGTH: 376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (223)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (250)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (259)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (273)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (361)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (364)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (367)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (372)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-302-409

Query Match 68.0%; Score 17; DB 9; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GAGGGCCAGAGGGCACA 21
Db 143 GAGGGCCAGAGGGCACA 127

RESULT 29

US-09-925-302-409/c
; Sequence 409, Application US/09925302
; Patent No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 409
; LENGTH: 376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (223)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (250)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (259)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (273)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (361)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (364)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (367)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (372)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-302-409

Query Match 68.0%; Score 17; DB 13; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GAGGGCCAGAGGGCACA 21
Db 143 GAGGGCCAGAGGGCACA 127

RESULT 30

US-09-918-995-6988/c
; Sequence 6988, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6988
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-918-995-6988

Query Match 68.0%; Score 17; DB 10; Length 406;
 Best Local Similarity 100.0%; Pred.No. 4e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GAGGCCACAGAGGGCACA 21
 |||||
 Db 73 GAGGCCACAGAGGGCACA 57

Search completed: June 20, 2004, 17:46:38
 Job time : 151.402 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 814.603 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgagaaacaggagtagcaccatgag 30

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_to.*

26: em_ats.*

27: em_un.*

28: em_vl.*

29: em_htg_hum.*

30: em_htg_inv.*

31: em_htg_inv.*

32: em_htg_inv.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rdd.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_hgo_hum.*

40: em_hgo_mus.*

41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|-----------|-------------|
| C 1 | 30 | 100.0 | 15619 | 1 | AE006916 | Mycobacte |
| C 2 | 30 | 100.0 | 38721 | 1 | MSGY219 | AD000013 |
| C 3 | 30 | 100.0 | 341957 | 15 | BX842572 | BX842572 |
| C 4 | 30 | 100.0 | 343050 | 1 | BX248334 | BX248334 |
| C 5 | 23.8 | 79.3 | 37821 | 1 | MLCB1770 | Z70722 |
| C 6 | 23.8 | 79.3 | 344050 | 1 | MLCBPRTN1 | AL583917 |
| C 7 | 21.2 | 70.7 | 98512 | 9 | AC022539 | AC022539 |
| C 8 | 21.2 | 70.7 | 154359 | 9 | AC021178 | AC021178 |
| C 9 | 21.2 | 70.7 | 158559 | 9 | AC091772 | AC091772 |
| C 10 | 21.2 | 70.7 | 158568 | 9 | AC069541 | AC069541 |
| C 11 | 21 | 70.7 | 188591 | 2 | AC118678 | AC118678 |
| C 12 | 21 | 70.7 | 208342 | 2 | AC105077 | AC105077 |
| C 13 | 20.6 | 68.7 | 510 | 6 | AR413807 | AR413807 |
| C 14 | 20.6 | 68.7 | 510 | 6 | BD109360 | BD109360 |
| C 15 | 20.6 | 68.7 | 1476 | 6 | E10616 | E10616 |
| C 16 | 20.6 | 68.7 | 1476 | 6 | E10861 | E10861 |
| C 17 | 20.6 | 68.7 | 1476 | 6 | AR399329 | AR399329 |
| C 18 | 20.6 | 68.7 | 1520 | 9 | AF182277 | AF182277 |
| C 19 | 20.6 | 68.7 | 2000 | 9 | AF081569 | AF081569 |
| C 20 | 20.6 | 68.7 | 3045 | 6 | AX018351 | AX018351 |
| C 21 | 20.6 | 68.7 | 3045 | 6 | AX018627 | AX018627 |
| C 22 | 20.6 | 68.7 | 3045 | 6 | AX332736 | AX332736 |
| C 23 | 20.6 | 68.7 | 3045 | 6 | AX409672 | AX409672 |
| C 24 | 20.6 | 68.7 | 3045 | 6 | BD136285 | BD136285 |
| C 25 | 20.6 | 68.7 | 3045 | 9 | HUMCP2BB | M29874 |
| C 26 | 20.6 | 68.7 | 39944 | 9 | AC011541 | AC011541 |
| C 27 | 20.6 | 68.7 | 42547 | 9 | AC023172 | AC023172 |
| C 28 | 20.6 | 68.7 | 150962 | 9 | AC092374 | AC092374 |
| C 29 | 20.6 | 68.7 | 164831 | 2 | AC145155 | AC145155 |
| C 30 | 20.6 | 68.7 | 222809 | 10 | AC117614 | AC117614 |
| C 31 | 20.4 | 68.0 | 41 | 6 | AX214506 | AX214506 |
| C 32 | 20.4 | 68.0 | 1595 | 4 | RABPEA | AX214506 |
| C 33 | 20.4 | 68.0 | 33385 | 2 | AC013058 | AC013058 |
| C 34 | 20.4 | 68.0 | 160294 | 2 | AC025621 | AC025621 |
| C 35 | 20.4 | 68.0 | 165654 | 2 | AC027747 | AC027747 |
| C 36 | 20.4 | 68.0 | 165928 | 3 | AC023725 | AC023725 |
| C 37 | 20.4 | 68.0 | 168208 | 2 | AC024359 | AC024359 |
| C 38 | 20.4 | 68.0 | 172784 | 3 | AC105352 | AC105352 |
| C 39 | 20.4 | 68.0 | 176921 | 9 | AC016080 | AC016080 |
| C 40 | 20.4 | 68.0 | 178142 | 9 | AC013546 | AC013546 |
| C 41 | 20.4 | 68.0 | 193572 | 10 | AC140268 | AC140268 |
| C 42 | 20.4 | 68.0 | 200363 | 9 | AC010895 | AC010895 |
| C 43 | 20.4 | 68.0 | 214271 | 2 | AC090737 | AC090737 |
| C 44 | 20.4 | 68.0 | 216834 | 10 | AC123867 | AC123867 |
| C 45 | 20.4 | 68.0 | 220715 | 9 | AC012410 | AC012410 |
| C 46 | 20.4 | 68.0 | 230614 | 10 | AL626772 | AL626772 |
| C 47 | 20.4 | 68.0 | 233913 | 2 | AC124048 | AC124048 |
| C 48 | 20.4 | 68.0 | 246196 | 2 | AC095869 | AC095869 |
| C 49 | 20.4 | 68.0 | 249044 | 2 | AC103086 | AC103086 |
| C 50 | 20.4 | 68.0 | 295177 | 3 | AE003430 | AE003430 |
| C 51 | 20.2 | 67.3 | 76681 | 9 | AL592223 | AL592223 |
| C 52 | 20.2 | 67.3 | 163183 | 9 | AC016574 | AC016574 |
| C 53 | 20.2 | 67.3 | 176447 | 2 | AC024894 | AC024894 |
| C 54 | 20.2 | 67.3 | 179648 | 10 | AL672122 | AL672122 |
| C 55 | 20.2 | 67.3 | 203134 | 9 | AC021077 | AC021077 |
| C 56 | 20.2 | 67.3 | 348050 | 1 | AP003581 | AP003581 |
| C 57 | 20 | 66.7 | 1588 | 8 | AY051044 | AY051044 |
| C 58 | 20 | 66.7 | 2374 | 8 | AF360288 | AF360288 |
| C 59 | 20 | 66.7 | 53000 | 2 | AC003656 | AC003656 |
| C 60 | 20 | 66.7 | 74350 | 8 | AP002543 | AP002543 |
| C 61 | 20 | 66.7 | 77508 | 2 | AC023005 | AC023005 |
| C 62 | 20 | 66.7 | 110000 | 2 | AC003656 | AC003656 |
| C 63 | 20 | 66.7 | 120871 | 9 | AP001054 | AP001054 |
| C 64 | 20 | 66.7 | 122638 | 9 | AB001523 | AB001523 |
| C 65 | 20 | 66.7 | 168277 | 2 | AC144598 | AC144598 |

| | | | | | | | | | | | | | |
|-------|------|------|--------|----|-----------|---------------------|-------|------|------|--------|----|-----------|---------------------|
| C 66 | 20 | 66.7 | 174503 | 9 | AC005697 | AC005697 Homo sapi | C 139 | 19.4 | 64.7 | 188928 | 10 | AC121814 | AC121814 Mus muscu |
| C 67 | 20 | 66.7 | 177906 | 2 | AC102047 | AC102047 Mus muscu | C 140 | 19.4 | 64.7 | 190173 | 10 | AC110573 | AC110573 Homo sapi |
| C 68 | 20 | 66.7 | 177917 | 2 | AC074236 | AC074236 Homo sapi | C 141 | 19.4 | 64.7 | 191154 | 2 | AC132150 | AC132150 Mus muscu |
| C 69 | 20 | 66.7 | 183890 | 9 | AC090868 | AC090868 Homo sapi | C 142 | 19.4 | 64.7 | 193694 | 2 | AC120728 | AC120728 Rattus no |
| C 70 | 20 | 66.7 | 188561 | 2 | AC115738 | AC115738 Mus muscu | C 143 | 19.4 | 64.7 | 196784 | 2 | AC113006 | AC113006 Mus muscu |
| C 71 | 20 | 66.7 | 194101 | 2 | AC090910 | AC090910 Homo sapi | C 144 | 19.4 | 64.7 | 198093 | 9 | AC018768 | AC018768 Homo sapi |
| C 72 | 20 | 66.7 | 220888 | 2 | AC136582 | AC136582 Rattus no | C 145 | 19.4 | 64.7 | 199113 | 2 | AC120411 | AC120411 Mus muscu |
| C 73 | 20 | 66.7 | 239292 | 2 | AC108329 | AC108329 Rattus no | C 146 | 19.4 | 64.7 | 200711 | 2 | AC133960 | AC133960 Homo sapi |
| C 74 | 20 | 66.7 | 240971 | 2 | AC130253 | AC130253 Rattus no | C 147 | 19.4 | 64.7 | 215863 | 2 | AC136812 | AC136812 Rattus no |
| C 75 | 20 | 66.7 | 340000 | 9 | AP001753 | AP001753 Homo sapi | C 148 | 19.4 | 64.7 | 224446 | 2 | AC132531 | AC132531 Rattus no |
| C 76 | 19.8 | 66.0 | 3432 | 14 | TACNP | M20304 Tacaribe vi | C 149 | 19.4 | 64.7 | 227162 | 2 | AC115138 | AC115138 Rattus no |
| C 77 | 19.8 | 66.0 | 103019 | 5 | AC091292 | AC091292 Takifugu | C 150 | 19.4 | 64.7 | 227976 | 2 | AC134029 | AC134029 Rattus no |
| C 78 | 19.8 | 66.0 | 215946 | 10 | AC124590 | AC124590 Mus muscu | C 151 | 19.4 | 64.7 | 228242 | 2 | AC109115 | AC109115 Rattus no |
| C 79 | 19.6 | 65.3 | 590 | 6 | AX184940 | AX184940 Sequence | C 152 | 19.4 | 64.7 | 238727 | 2 | AC114539 | AC114539 Mus muscu |
| C 80 | 19.6 | 65.3 | 649 | 6 | AX187808 | AX187808 Sequence | C 153 | 19.4 | 64.7 | 239467 | 2 | AC111384 | AC111384 Rattus no |
| C 81 | 19.6 | 65.3 | 695 | 11 | BV048790 | BV048790 S212P6227 | C 154 | 19.4 | 64.7 | 240315 | 2 | AC099474 | AC099474 Rattus no |
| C 82 | 19.6 | 65.3 | 778 | 6 | AX186743 | AX186743 Sequence | C 155 | 19.4 | 64.7 | 246402 | 2 | AC134800 | AC134800 Rattus no |
| C 83 | 19.6 | 65.3 | 785 | 10 | AF044280 | AF044280 Rattus no | C 156 | 19.4 | 64.7 | 248439 | 2 | AC098173 | AC098173 Rattus no |
| C 84 | 19.6 | 65.3 | 2569 | 9 | AC091287 | AC091287 Homo sapi | C 157 | 19.4 | 64.7 | 248537 | 2 | AC107098 | AC107098 Rattus no |
| C 85 | 19.6 | 65.3 | 4772 | 3 | BT003507 | BT003507 Drosophil | C 158 | 19.4 | 64.7 | 249230 | 2 | AC121466 | AC121466 Rattus no |
| C 86 | 19.6 | 65.3 | 6248 | 10 | AX1212196 | AX1212196 Mus muscu | C 159 | 19.4 | 64.7 | 253346 | 2 | AC128440 | AC128440 Rattus no |
| C 87 | 19.6 | 65.3 | 6847 | 9 | HSW806465 | BX538231 Homo sapi | C 160 | 19.4 | 64.7 | 255746 | 2 | AC122622 | AC122622 Rattus no |
| C 88 | 19.6 | 65.3 | 11191 | 2 | AC020352 | AC020352 Drosophil | C 161 | 19.4 | 64.7 | 256813 | 2 | AC135301 | AC135301 Rattus no |
| C 89 | 19.6 | 65.3 | 111107 | 9 | AC005924 | AC005924 Homo sapi | C 162 | 19.4 | 64.7 | 261538 | 2 | AC112004 | AC112004 Rattus no |
| C 90 | 19.6 | 65.3 | 145414 | 9 | HS78F24 | AL022336 Human DNA | C 163 | 19.4 | 64.7 | 272627 | 2 | AC129054 | AC129054 Rattus no |
| C 91 | 19.6 | 65.3 | 150587 | 3 | AC007549 | AC007549 Drosophil | C 164 | 19.4 | 64.7 | 316813 | 8 | SCCHR111 | SCCHR111 |
| C 92 | 19.6 | 65.3 | 162522 | 2 | AC021014 | AL529193 Zebrafish | C 165 | 19.2 | 64.0 | 912 | 8 | AF106929 | AF106929 Medicago |
| C 93 | 19.6 | 65.3 | 169816 | 2 | AC118600 | AC118600 Homo sapi | C 166 | 19.2 | 64.0 | 1136 | 8 | D89219 | D89219 Schizosacch |
| C 94 | 19.6 | 65.3 | 175326 | 9 | CNS01DTQ | AL132838 Human chr | C 167 | 19.2 | 64.0 | 1210 | 8 | TOBFESA | TOBFESA |
| C 95 | 19.6 | 65.3 | 181771 | 3 | AC008340 | AC008340 Drosophil | C 168 | 19.2 | 64.0 | 1286 | 8 | TOBPESC | TOBPESC |
| C 96 | 19.6 | 65.3 | 198100 | 2 | BX897683 | BX897683 Mus muscu | C 169 | 19.2 | 64.0 | 1598 | 10 | BC010459 | BC010459 Mus muscu |
| C 97 | 19.6 | 65.3 | 198772 | 10 | AL606906 | AL606906 Mouse DNA | C 170 | 19.2 | 64.0 | 1655 | 10 | MWNEFA | MWNEFA |
| C 98 | 19.6 | 65.3 | 205268 | 2 | BX890581 | BX890581 Danio rer | C 171 | 19.2 | 64.0 | 2250 | 8 | AB012387 | AB012387 Schizosac |
| C 99 | 19.6 | 65.3 | 209197 | 9 | AC007389 | AC007389 Homo sapi | C 172 | 19.2 | 64.0 | 2870 | 3 | AF029249 | AF029249 Mytilus e |
| C 100 | 19.6 | 65.3 | 209197 | 9 | AF288742 | AF288742 Homo sapi | C 173 | 19.2 | 64.0 | 3928 | 8 | AF000150 | AF000150 Schizosac |
| C 101 | 19.6 | 65.3 | 221548 | 2 | AC099714 | AC099714 Mus muscu | C 174 | 19.2 | 64.0 | 4039 | 10 | AF081499 | AF081499 Mus muscu |
| C 102 | 19.6 | 65.3 | 237178 | 2 | AC105297 | AC105297 Mus muscu | C 175 | 19.2 | 64.0 | 10029 | 1 | AE012085 | AE012085 Xanthomon |
| C 103 | 19.6 | 65.3 | 242649 | 1 | AE0108648 | AE0108648 Rattus no | C 176 | 19.2 | 64.0 | 13630 | 1 | AE007459 | AE007459 Streptoco |
| C 104 | 19.6 | 65.3 | 300956 | 1 | AE016963 | AE016963 Coccidia | C 177 | 19.2 | 64.0 | 14089 | 8 | SPAPB18E9 | SPAPB18E9 |
| C 105 | 19.6 | 65.3 | 348873 | 3 | AE003790 | AE003790 Drosophil | C 178 | 19.2 | 64.0 | 21706 | 6 | AD218804 | AD218804 Sequence |
| C 106 | 19.6 | 65.3 | 655 | 11 | BV018096 | BV018096 S208P653 | C 179 | 19.2 | 64.0 | 21706 | 6 | BD003716 | BD003716 Polynucle |
| C 107 | 19.4 | 64.7 | 1260 | 14 | TRVAPGD | L34031 Turkey rhin | C 180 | 19.2 | 64.0 | 39781 | 8 | SPCC1739 | SPCC1739 |
| C 108 | 19.4 | 64.7 | 1260 | 14 | TRVAPGD | L34033 Turkey rhin | C 181 | 19.2 | 64.0 | 110000 | 2 | LMFLCHR32 | LMFLCHR32 |
| C 109 | 19.4 | 64.7 | 1260 | 14 | TRVAPGD | L34034 Turkey rhin | C 182 | 19.2 | 64.0 | 158764 | 10 | AC132083 | AC132083 Homo sapi |
| C 110 | 19.4 | 64.7 | 1485 | 3 | PFACSNURI | M11031 Plasmodium | C 183 | 19.2 | 64.0 | 158952 | 2 | AC096745 | AC096745 Homo sapi |
| C 111 | 19.4 | 64.7 | 1587 | 10 | BC049142 | BC049142 Mus muscu | C 184 | 19.2 | 64.0 | 172625 | 2 | AC124108 | AC124108 Mus muscu |
| C 112 | 19.4 | 64.7 | 1587 | 10 | BC049142 | BC049142 Mus muscu | C 185 | 19.2 | 64.0 | 172625 | 2 | AC144639 | AC144639 Ornithorh |
| C 113 | 19.4 | 64.7 | 2261 | 10 | BC058745 | BC058745 Mus muscu | C 186 | 19.2 | 64.0 | 185695 | 2 | AC107837 | AC107837 Mus muscu |
| C 114 | 19.4 | 64.7 | 3337 | 8 | AC065496 | AC065496 Oryza sat | C 187 | 19.2 | 64.0 | 210172 | 2 | AC127686 | AC127686 Mus muscu |
| C 115 | 19.4 | 64.7 | 3465 | 8 | SCABPI | XS1780 Yeast ABPI | C 188 | 19.2 | 64.0 | 248489 | 2 | AC115656 | AC115656 Rattus no |
| C 116 | 19.4 | 64.7 | 51420 | 9 | AL139412 | AL139412 Human DNA | C 189 | 19.2 | 64.0 | 249625 | 2 | AC094817 | AC094817 Rattus no |
| C 117 | 19.4 | 64.7 | 95845 | 8 | AF411807 | AF411807 Lycopersi | C 190 | 19.2 | 64.0 | 254098 | 2 | AC096236 | AC096236 Rattus no |
| C 118 | 19.4 | 64.7 | 96180 | 9 | AC005868 | AC005868 Homo sapi | C 191 | 19.2 | 64.0 | 262405 | 2 | AC094802 | AC094802 Rattus no |
| C 119 | 19.4 | 64.7 | 99003 | 2 | AL390756 | AL390756 Homo sapi | C 192 | 19.2 | 64.0 | 266051 | 2 | AC126711 | AC126711 Rattus no |
| C 120 | 19.4 | 64.7 | 120770 | 9 | AL513190 | AL513190 Human DNA | C 193 | 19.2 | 64.0 | 274906 | 2 | AC094962 | AC094962 Rattus no |
| C 121 | 19.4 | 64.7 | 125750 | 2 | AC092011 | AC092011 Felis cat | C 194 | 19.2 | 64.0 | 349380 | 6 | AX571764 | AX571764 Sequence |
| C 122 | 19.4 | 64.7 | 141653 | 10 | AC089919 | AC089919 Mus muscu | C 195 | 19.2 | 63.3 | 280 | 6 | AX904472 | AX904472 Sequence |
| C 123 | 19.4 | 64.7 | 151700 | 9 | AC133528 | AC133528 Homo sapi | C 196 | 19.2 | 63.3 | 280 | 6 | BSA299740 | BSA299740 Homo sapi |
| C 124 | 19.4 | 64.7 | 152299 | 2 | AC125012 | AC125012 Mus muscu | C 197 | 19.2 | 63.3 | 1167 | 9 | BC005050 | BC005050 Homo sapi |
| C 125 | 19.4 | 64.7 | 153675 | 2 | AC092012 | AC092012 Felis cat | C 198 | 19.2 | 63.3 | 1578 | 9 | BC005050 | BC005050 Homo sapi |
| C 126 | 19.4 | 64.7 | 153675 | 2 | AC139188 | AC139188 Rattus no | C 199 | 19.2 | 63.3 | 2019 | 9 | BC050853 | BC050853 Homo sapi |
| C 127 | 19.4 | 64.7 | 159857 | 9 | AC018767 | AC018767 Homo sapi | C 200 | 19.2 | 63.3 | 2128 | 4 | AF074085 | AF074085 Felis cat |
| C 128 | 19.4 | 64.7 | 159857 | 10 | AL551075 | AL551075 Mouse DNA | C 201 | 19.2 | 63.3 | 2318 | 9 | AF538150 | AF538150 Homo sapi |
| C 129 | 19.4 | 64.7 | 160545 | 2 | AP003508 | AP003508 Oryza sat | C 202 | 19.2 | 63.3 | 2377 | 9 | AF538150 | AF538150 Homo sapi |
| C 130 | 19.4 | 64.7 | 160557 | 2 | AC021795 | AC021795 Homo sapi | C 203 | 19.2 | 63.3 | 31004 | 9 | HGN1366 | HGN1366 |
| C 131 | 19.4 | 64.7 | 163136 | 2 | AP005913 | AP005913 Oryza sat | C 204 | 19.2 | 63.3 | 32217 | 9 | AC017265 | AC017265 Drosophil |
| C 132 | 19.4 | 64.7 | 175727 | 10 | AC127580 | AC127580 Mus muscu | C 205 | 19.2 | 63.3 | 36993 | 2 | AC018007 | AC018007 Drosophil |
| C 133 | 19.4 | 64.7 | 179390 | 10 | AC122190 | AC122190 Mus muscu | C 206 | 19.2 | 63.3 | 39468 | 9 | AC007205 | AC007205 Homo sapi |
| C 134 | 19.4 | 64.7 | 181763 | 2 | AC136735 | AC136735 Mus muscu | C 207 | 19.2 | 63.3 | 42224 | 1 | MUCB33 | MUCB33 |
| C 135 | 19.4 | 64.7 | 181763 | 2 | AC132728 | AC132728 Rattus no | C 208 | 19.2 | 63.3 | 67671 | 2 | AC102113 | AC102113 Mus muscu |
| C 136 | 19.4 | 64.7 | 183008 | 2 | AC019276 | AC019276 Homo sapi | C 209 | 19.2 | 63.3 | 69334 | 2 | AC101350 | AC101350 Mus muscu |
| C 137 | 19.4 | 64.7 | 185571 | 9 | AC133781 | AC133781 Homo sapi | C 210 | 19.2 | 63.3 | 106866 | 10 | AL732550 | AL732550 Mouse DNA |
| C 138 | 19.4 | 64.7 | 185921 | 2 | AC132645 | AC132645 Rattus no | C 211 | 19.2 | 63.3 | 110000 | 2 | AC096436 | AC096436 Rattus no |

TLKYKVVYRFGLYGVEEYHPDLQIVGYADPVRMEHPAQAGPLDY"

gene

CDS

/gene="WT0016"
/note="similar to PID:1262359; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamine amidotransferase, class I"
/protein_id="AAK44238.1"
/db_xref="GI:13879057"

/translation="NRILVNDYDSFVNLVQYLGQIEAEVNRNDHRLSDAAVA
GQFDVLLSPGPGTPERAGASVLIHACAAHETPLGVCVGHQAGIAGFAGTVDRAPE
LLHGKTSVFTNNGVQLGDLDPFTATRYHSLTILPKSLPAVLRTVARTSGVIMAVQ
HTGLPIGVQFHPHSILTEGGHRIILANWLTCGWTQDDTLVRLENEVLTAISPHFT
STASAGEATGRTSA"
complement (1579..3459)
/gene="WT0017"
complement (1579..3459)
/gene="WT0017"
/note="similar to PID:2344821; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine/threonine protein kinase"
/protein_id="AAK44239.1"
/db_xref="GI:13879058"

gene

CDS

/translation="MTTSHLSDRYEIGLPGCGMEVHLARLRLHRDVAVKVLRA
DLADPSYLRPREAQAALNHPAIVAVYDTGEAETPAGPLPIVWEYVDGVLDRD
IVHTEGPTPKRAIEVIADACALNFHQNGIHRDKVPANIMISATNVCVMDFGIA
RAIADGNSVTQTAIVTAQYLSPEQARGSDVARDVSLGCVLYEVLTPGPPFTG
DSPSVAYQVREDDIPPSARHGLSDLDVAVLKAKXPNENRYQTAAMRADLVRV
HNGEPPEAPKVLTAERTSLSSAAGNSLGPRTDPLPROLDDTDDRSISGVGRWA
VAVLAVLVVTVTAINTFGGITRDVQVQVDRGSSADATATLQNRGFKIRTQKPS
TIPDHVLTGTPAANTSAGDEITVNSVTPGOREIPDVSTLIYAEAVKRLTAAGFG
RFKQNSVTELPVKVIGTNPANQTSAINTVIIIVGSGPATKIDPVAGQTVDA
QKMLNVGTFKFSQVSDSPRAGEVTGNPAGTTVPVDSVIBLQSKNGQVFPMDL
SGMFWDAEPRLALGWTMLDKGADVAGSGQENRVVYQNPAGTGVNRDGIITLRF
GQ"

gene

CDS

/gene="WT0018"
complement (3456..4751)
/gene="WT0018"
/note="similar to PID:2344822; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine/threonine protein kinase"
/protein_id="AAK44240.1"
/db_xref="GI:13879059"

/translation="MSPRVGVTLSGRYLRQLRIATGGMQGVWEAVDNLGRVAVKVL
KSEFSDPEFTFRFAERTTAMLNHPGIAHVHDYQMGNGERTAYLWELVNGEP
LNSVLRKTRLSLHLDMLQETGALQIAHAGLVHRDVKVPKNILITPTGVQVKITDF
GAKAVDAAPYQTVGMVTAQYIAPEQALGHDPASDVYSLGVVGYEAVSGKRPPA
GDGLATVMKHIKSPPPPLDLPNNVRELIEITLVKNPAMRYSGSGPFADAVAVRA
GRPPRSPPTPPGRAPALIPSGTTARVANSAGRTAASRRNPATCGHRPPRTFS
SQGALUWAAGVIGALAIITIAVLVIKAPGNSPQAPPTPTVTITGNPPASNTGGTDA
SPRLNWERGTRHSGLSQVSPVPTPHSRASLARYETAQ"

gene

CDS

/complement (4748..6223)
/gene="WT0019"
complement (4748..6223)
/gene="WT0019"
/note="similar to GP:5139605; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein, putative"
/protein_id="AAK44241.1"
/db_xref="GI:13879060"

gene

CDS

/translation="MNASLRISVYTMALIVLLLNATMQVTADGLADPNQRVL
LDESQRQGITAGQLLAVSDGFRFLRVLPNEVYAPVTFGTSLSYSLTALER
AEDPILNGSDRLFRGLADFTGRDPRGNVDVTINPRITQAGWAMQOQCGYCKG
AAVLEPSTGKILALVSPSYDNPILLASHNFEVQAQWRLGDNFASPLTNRASISY
PGSTFKVITTAALAAAGATETEQTAAPTIPLPGSTAGLENYGAPCDEFTSLRE
AFVKSCNTAFVQIGRTGADALRSARFAPVTPLOVAESTVGPIDPSAALGM
TSIQGQVALTFLANAEIATANGITMRPYLVGSLKGPDLANI-STTVGQQRRAVS
POVAKILTELMVGAKEVQKGAIPQVQIASKTGTAEHGTDPRTPPHAWYIAFAPAQ
APKAVAVLIVENGADRLSATGCALAAPIGRAVIEAALQGEF"
complement (6220..7629)
/gene="WT0020"
complement (6220..7629)
/gene="WT0020"
/note="similar to GB:D10483 SP:P16457 GB:M30807 PID:146039
PID:216503; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsW"
/protein_id="AAK44242.1"
/db_xref="GI:13879061"

gene

CDS

/translation="MTTQLQAPVAVTPELPTRRNAELLLCFAAVITFAALLVQANQ
DOGVPMDLTSYGLAFLTLFGSAHLAIRFPAPVTDPLLPVVALINGLGLVIMHRLDLV
DNEIGEHHPHSANQOMLTGVAFAFVAVTFKDRHOLARYGVICGLAGLVFLAVPA
LLPAALSQNGAKIMIRLPGFSIQPAEFSKILLIFESAVALVAKEGLETSAGKHLGM
TLPRDLAPLAAWVISGVNVFEDKGLCASLLITSLVVVYLATQFSWVIGLTL
FAAGTLVAFIFEHVRLRVQMTLDFADPGTGYQIVQSLFSFATGIFGGLNGQP
DTVPAASDTFIIAFAFGEELGVLTLAILMTYIVIRGLRTAINTDRSFGKLLAAGLS
STLAIQLFTIVGVGTRLLIPLTGLTTPWMSYGGSSLLANYILLAILARISHGARREPT
RPNKSPITAACTEVIERY"
complement (7626..9161)
/gene="WT0021"
complement (7626..9161)
/gene="WT0021"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44243.1"
/db_xref="GI:13879062"

gene

CDS

/translation="MTLVLRVAASDRGLVRANNEDSVYAGARLLALADMGHGAAGE
VASQVIAALHLDDEPGDOLLAKLDAARAGNSAIAAQVEMEPDLEGMGTTLTAL
FAGNLVILHIGDSRYLLRDELQITKDDTFTVQLTVEGRITPEEASHSPQSLIM
KALGHEVETLTMREARAGRYLLCSDGLSDPVSDETLBALQIPEVAESAHRLIEL
ALRGGGPNVTWVADVVDYDYGOTQPILAGVSGDDDLTLPTNPAAGSAISORKE
IVRVVPPQADTFESPESGSRBLAFVAVLTVMTAGLIGALIRSNVYVADYAGSVS
INRGIGQSLGMSLQPYLMGCLSPRNELSIQISQSGGPDLDCHLMKLEDLRPERAQ
VWAPAGLTDDAICQLRELAANSLPPCPAPRATSPGPRPAPPTTSETTBNVTSSP
ASPSPTTSASAPTGTTPAITSASPAASPPTPMTPTTSPTMAALPPPPPPQGDICR
AAA"
complement (9259..9726)
/gene="WT0022"
complement (9259..9726)
/gene="WT0022"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44244.1"
/db_xref="GI:13879063"

gene

CDS

/translation="MQGLVQLTRAGFLMLLVFWISVLRILKTDIVAPTAVAMVRG
LALRGTLLQQRRAHARLVVTEGALTGARITISEQVLLGRADDSLTVDITDFAST
RHARLUSMGSEWYEDLGSTNGTGYLDRAKVTTAVRVPTGTFTVRIKTAIELRP"
complement (9850..11415)
/gene="WT0023"
complement (9850..11415)
/gene="WT0023"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44245.1"
/db_xref="GI:13879064"

gene

CDS

/translation="MGSOKRLVORVERKLEOTVGDAFARIFGSGISVPOVEALLRREA
ADGTSLOGNRLAPNEVIIITLVHDFEKLADPELAKSTGEARLADYIQOQWYTC
DVVRFQSSNLHTQFARGVTPDVEHPTHEPDIARQPSNHAFAEGFAPVAPMSDMS
SYRGQGGQGRDEYDDRYARPDPRGGPDQGGSDPRGGYPPETGGYPPQGYPRP

h R., Parkhill J., Garra
meier K., Cas S., Barr
D., Chillingworth T.,
tles S., Hamlin N., Hol
J., Moule S., Murphy J.
ndream M.A., Rogers J.
es R., Sulston J.E., Ta
biology of Mycobacter
44 (1998) .

r M.J., Medigue C., Col
f the genome sequence c
:2967-2973 (2002).

Location/Qualifiers

1. 341957
/db_xref=taxon:83332"
/mol_type="genomic DNA"
/organism="Mycobacteri
/strain="H37Rv"

1. 1524
/evidence=EXPERIMENTAL
/note="rv00001, (MU0001
dnaA, chromosomal repl
citations below), equiva
CHROMOSOMAL REPLICATIO
Q46388 [DNAA_MYCFA from
Q917L7 [DNAA_MYCPA from
aal; P49390 [DNAA_MYCAV
P49992 [DNAA_MYCSM from
etc. Also highly simil
e.g. Q9ZHT5 [DNAA_STRCH
PROTEIN from Streptomy
Q9ZH76 [DNAA_SPRRE from
DNAA_ECOLI|F031004|B37070
protein from Escherich
overlap; opt. Contain
coversl; etc. Contain
A A (P-loop) and P901008
THE DNAA FAMILY. Note
been taken as base 1 o
H37Rv genomic sequenc
/trans_table=11
/function=P_LAYS AN IM
REGULATION OF CHROMO
OF REPLICATION; IT BIND
AT A 9 BP CONSENSUS ID
DNAA BINDS TO ATP AND
BINDS THE ORIGIN OF REP
EXHIBITED WEAK ALPase
/locus_tag="RV0001"
/gene="dnaA"

ier T., Churcher C., Harris D.,
Y III C.E., Tekala F., Badcock K.,
Connor R., Davies R., Devlin K.,
royd S., Hornaby T., Jageis K.,
... Oliver S., Osborne J.,
Rutter S., Seeger K., Skelton S.,
aylor K., Whitehead S., Barrell B.G.;
um tuberculosis from the complete
e S.T.;
f Mycobacterium tuberculosis H37Rv";
Bank/DBJ databases.
um tuberculosis sequencing and
Trust Genome Campus, Hinxton,
Moleculaire Bacterienne, Institut
Paris Cedex 15, France E-mail:
at the Sanger Centre
'M_tuberculosis/';
um tuberculosis H37Rv"
... , MV029 01, P49993), len: 507 aa.
ocation initiator protein (see
valent to other Mycobacterial
N INITIATOR PROTEINS e.g.
Mycobacterium leprae (502 aa);
Mycobacterium paratuberculosis (509
from Mycobacterium avium (508 aa);
Mycobacterium smegmatis (504 aa);
ear to others except in N-terminus
CHROMOSOMAL REPLICATION INITIATOR
ces chrysomallus (624 aa);
2 streptomycetes reticuli (643 aa);
ia coli strain K12 (467 aa), FASTA
o, (43.2% identity in 389 aa
s P00017 ATP/GTP-binding site motif
DNAa protein signature. BELONGS TO
that the first base of this gene has
f the Mycobacterium tuberculosis
".
PORTANT ROLE IN THE INITIATION AND
MAL REPLICATION. BINDS TO THE ORIGIN
DS SPECIFICALLY DOUBLE-STRANDED DNA
(DNAa BOX): 5'-TTATC(C/A)(C/A)-3'.
TO ACIDIC PHOSPHOLIPIDS. DNAa PROTEIN
PLICATION (oric), ATP AND ADP, AND
ACTIVITY."

Hinxton, Cambridge CB10 1SA, UK. P74 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France.

Location/Qualifiers

1. 343050
 /organism="Mycobacterium bovis subsp. bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 1. 1524
 /locus_tag="dnaA"
 /locus_tag="Mb0001"
 1. 1524
 /gene="dnaA"
 /locus_tag="Mb0001"
 /notes="Mb0001, dnaA, len: 507 aa. Equivalent to Rv0001, len: 507 aa, from Mycobacterium tuberculosis strain H37Rv, (99.6% identity in 507 aa overlap). dnaA, chromosomal replication initiator protein (see citations below), equivalent to other Mycobacterial CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g. P46388|DNAA_MYCLE from Mycobacterium leprae (502 aa); Q9L7L7|DNAA_MYCPA from Mycobacterium paratuberculosis (509 aa); P49990|DNAA_MYCAV from Mycobacterium avium (508 aa); P49992|DNAA_MYCSM from Mycobacterium smegmatis (504 aa); etc. Also highly similar to others except in N-terminus e.g. Q92H75|DNAA_STRCH CHROMOSOMAL REPLICATION INITIATOR PROTEIN from Streptomyces chrysosallus (624 aa); Q92H76|DNAA_STRRE from Streptomyces reticuli (643 aa); DNAA_ECOLI P03004|B3702 chromosomal replication initiator protein from Escherichia coli strain K12 (467 aa), FASTA scores: opt: 966, E(): 0, (43.2% identity in 389 aa overlap); etc. Contains PS00017 ATP/GTP-binding site motif A (p-loop) and PS01008 DNAa protein signature. BELONGS TO THE DNAa FAMILY. Note that the first base of this gene has been taken as base 1 of the Mycobacterium bovis genomic sequence."
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAa"
 /protein_id="CAD92863.1"
 /db_xref="GI:31616763"
 /translation="MTDPPGSGFTTVNAVVELNGPKVDDGPPSDANLSAPLTQQ
 RAWNLVQPTIVGFAALLVPFSFVQNEIERHLRAFTIDLRRLRHQIQLGVRAP
 PATDEADTTVPSENATTPDTTNDDEIDDSAAAGDNQHSWPSYFPTPRNTDS
 ATAGVTGLNRYTDTVIGASNFRAHAALAAEPARAYNPLFTWGESGLGKTHLL
 HAAGYACRLPFGVRVYVSTNFTNDFINSLDRDKVAFKGRVDDVLLVDIDQPI
 EGKGIQEEPHFTHTLNANKQIVISDPPKOLATLERLRTREPWGLITDVPQE
 LETRIALRKQMERLAIEDDVLELIASSIERNIERNLEGLALRVTFASFNLTPDK
 ALAEIVRLDIADANTQIISAATMAATAEFTDTTVEELRGPKTRALAGRIAMYL
 CRELTDLSPKIGQAFGRDHTTVMYAQKILSENAERREVFHVKELTIRQRKR"
 2052. 3260
 /gene="dnaA"
 /locus_tag="Mb0002"
 /locus_tag="Mb0002"
 2052. 3260
 /gene="dnaA"
 /locus_tag="Mb0002"
 /EC_number="2.7.7.7"
 /note="Mb0002, dnaN, len: 402 aa. Equivalent to Rv0002, len: 402 aa, from Mycobacterium tuberculosis strain H37Rv, (99.8% identity in 402 aa overlap). dnaN, DNA polymerase III (beta chain) (EC 2.7.7.7) (see citations below), equivalent to other Mycobacterial DNA POLYMERASES III BETA CHAIN e.g. NP_301130.1|NC_002677 from Mycobacterium leprae (399 aa); Q9L7L6|DP3B_MYCPA from Mycobacterium avium subsp. paratuberculosis (399 aa); P52851|DP3B_MYCSM from Mycobacterium smegmatis (397 aa); etc. Also highly similar to others e.g. P27903|DP3B_STRCO DNA POLYMERASE III BETA CHAIN from Streptomyces coelicolor (376 aa), FASTA scores: opt: 1189, E(): 0, (52.8% identity in 337 aa overlap); P21174|DP3B_MICLU from Micrococcus luteus (310 aa); P52023|DP3B_SYNPT from Synecchococcus sp. strain PCC 7942

(375 aa); etc. Overlaps and extends CDS in neighbouring cosmid MICV10H4.01."
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA NUCLEOTIDYLTRANSFERASE)"
 /protein_id="CAD92864.1"
 /db_xref="GI:31616764"
 /translation="WDAATTRVGLTDLTFLRLRESFADAVSWAKNLPARPAVPLSG
 VLLTGSNGLTISGDFEVSAAEAGVSPGSLVSGRLLSITRALPNKPVGVHVH
 EGNVALTCGNARSLFPMVEDYPTLTLEETGLLPAELFAEASIVATAAGRDDIT
 LPMLTGIRVEILGETVIVLAADRFLRAVRELKWSASSPDIEAAVLVPKATILAEAKAG
 IGGSDVRLSLGTGPGKGLLGISGNGKRSITRLDRAEFKFKQLLPTHTVAITMD
 VAEITIAIKLVALVADRGAVRMEFAGSVRLSAGADDVGRAEEDLVVDYAGEPLTIA
 FNPTVLTDLGLSSLSRVSFGFTTAGKPAALLRPVSGDDRPVAGLNGNGPFFAVSTDYV
 YLLMEVRLPG"
 3280. 4437
 /gene="recF"
 /locus_tag="Mb0003"
 3280. 4437
 /gene="recF"
 /locus_tag="Mb0003"
 /note="Mb0003, recF, len: 385 aa. Equivalent to Rv0003, len: 385 aa, from Mycobacterium tuberculosis strain H37Rv, (99.5% identity in 385 aa overlap). recF, DNA replication and repair protein (see citations below), equivalent to others Mycobacterial DNA replication and repair proteins e.g. NP_301131.1|NC_002677 from Mycobacterium leprae (385 aa); Q9L7L5|RECF_MYCPA from Mycobacterium avium subsp. paratuberculosis (385 aa); P50916|RECF_MYCSM from Mycobacterium smegmatis (384 aa); etc. Also highly similar to others e.g. P36176|RECF_STRCO DNA REPLICATION AND REPAIR PROTEIN from Streptomyces coelicolor (373 aa); REPAIR PROTEIN from Streptococcus sp. strain PCC 6803 (384 aa); NP_469352.1|NC_003212 from Listeria innocua (370 aa); etc. Contains PS00017 ATP/GTP-binding site motif A (p-loop). PS00617 recF protein signature 1, and PS00618 recF protein signature 2. BELONGS TO THE RECF FAMILY."
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="DNA REPLICATION AND REPAIR PROTEIN RECF (SINGLE-STRAND DNA BINDING PROTEIN)"
 /protein_id="CAD92865.1"
 /db_xref="GI:31616765"
 /translation="MIVRHLDPRFRWACVDLELHPGRTVTVGNGYKTNLLIALW
 YETTLGSHVSADLEPLIRVGTDRIVSTIVNDRECAVDLEIATGRVMAKLNSSV
 RSTRDVGVIRAVLFAPEDLGLVRGDPADRRYLDLAIVRPAAVRAEYRVVRQ
 RTALLKSVPCARYGRDGVFTDLVMDSLAEHGAELVAARIDLVAQLAPEYKAYQL
 LAPERSASIGYRASMDVTGPSEQDTRQLLAARLLAALAAARDAERLGVCLVGPV
 RDLLRLRGDQPAKGFAHGAEWSLAVRLAAYOLLRVDDGPEVLLDDVFAELDM
 RRLAATAEAEQVLTVAVLEDPAGHDARVIDRADDTGSMVSLP"
 4434. 4997
 /locus_tag="Mb0004"
 4434. 4997
 /locus_tag="Mb0004"
 /note="Mb0004, recP, len: 187 aa. Equivalent to Rv0004, len: 187 aa, from Mycobacterium tuberculosis strain H37Rv, (99.5% identity in 187 aa overlap). Conserved hypothetical protein (see citation below), highly similar, but longer 21 aa in N-terminus, to AAF33696.1|AF222789 unknown protein from Mycobacterium avium subsp. paratuberculosis (166 aa); and highly similar to NP_301132.1|NC_002677 conserved hypothetical protein from Mycobacterium leprae (189 aa); S70990 hypothetical protein from Mycobacterium smegmatis (194 aa). Also similar to in C-terminus to C-terminal part of P35925|YREG_STRCO HYPOTHETICAL 19.8 KDA PROTEIN (IN RECF-GYR INTERGENIC REGION) from Streptomyces coelicolor (190 aa), FASTA scores: opt: 404, E(): 3.9e-18, (40.7% identity in 189 aa overlap)."
 /codon_start=1
 /transl_table=11

FEATURES

source

gene

CDS

gene

CDS

```

/product="CONSERVED HYPOTHETICAL PROTEIN"
/db_xref="GI:31616786"
/translation="MTGSDVRPDQNGRGLMKSPGLDLVRLTDEARAAARQCQDAG
RGVAVASGRVAGRRWSGFGDIRDPQLPKGAARELAKRGWSVRVAGMVLGOW
SAVVGQIAEHAHPTALNDGVLSTVAESTAWATQLRINQALLAKIAAAVGVRSRL
KITGPAAPSWRKGPRHAGRGPRDTYG"
5123..7267
/gene="gyrB"
CDS
5123..7267
/gene="gyrB"
/locus_tag="Mb00005"
/gene="gyrB"
/locus_tag="Mb00005"
/EC_number="5.99.1.3"
/notes="Mb00005. gyrB, len: 714 aa. Equivalent to Rv0005,
len: 714 aa, from Mycobacterium tuberculosis strain H37RV,
(99.9% identity in 714 aa overlap). gyrB, DNA gyrase
subunit B (EC 5.99.1.3) (see citations below), equivalent,
except in N-terminus, to other Mycobacterial DNA GYRASES
SUBUNIT B e.g. T10005 from Mycobacterium leprae (697 aa);
Q9L7L3|GYRB MYCPA from Mycobacterium avium subsp.
Paratuberculosis (677 aa) (has its N-terminus shorter);
F48355|GYRB MYCSM from Mycobacterium smegmatis (675 aa);
etc. Also highly similar to others e.g. T10969 from
Streptomyces coelicolor (686 aa); P50075|GYBS_STRSH from
Query Match 100.0%; Score 30; DB 1; Length 343050;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGCGAGAACGAGGATGACCAATGAG 30
Db 24985 TGGCGAGAACGAGGATGACCAATGAG 24956

RESULT 5
MCLB1770/c
LOCUS MCLB1770 37821 bp DNA linear BCT 29-AUG-1997
DEFINITION Mycobacterium leprae cosmid B1770.
ACCESSION 270722
VERSION 270722.1 GI:2344819
KEYWORDS Anthranilate synthase; DNA A; DNA gyrase; DnaN; Intein; PBP;
peptidyl-prolyl isomerase; pseudogene; RecF; RodA; Ser; Thr-protein
kinases; tRNA-Ala; tRNA-Ile; tRNA-Leu.
SOURCE Mycobacterium leprae
ORGANISM Mycobacterium leprae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE 1 (bases 1 to 37821)
Figsli.H., De Rossi.E., Salazar.L., Cantoni.R., Labo.M.,
Riccardi.G., Takiff.H.E., Eigmeier.K., Bergh.S. and Cole.S.T.
Gene arrangement and organization in a approximately 76 kb fragment
encompassing the oric region of the chromosome of Mycobacterium
leprae
Mol. Microbiol. 7 (2), 197-206 (1993)
JOURNAL 93188700
MEDLINE 8446027
PUBMED 8446027
REFERENCE 2 (bases 1 to 37821)
Figsli.H., De Rossi.E., Salazar.L., Cantoni.R., Labo.M.,
Riccardi.G., Takiff.H.E., Eigmeier.K., Bergh.S. and Cole.S.T.
Gene arrangement and organization in a approximately 76 kb fragment
encompassing the oric region of the chromosome of Mycobacterium
leprae
Mol. Microbiol. 7 (2), 197-206 (1993)
JOURNAL 93188700
MEDLINE 8446027
PUBMED 8446027
REFERENCE 3 (bases 1 to 37821)
Figsli.H., Vincent.V. and Cole.S.T.
Homing events in the gyrA gene of some mycobacteria
Proc. Natl. Acad. Sci. U.S.A. 93 (8), 3410-3415 (1996)
JOURNAL 96194983
MEDLINE 8622943
PUBMED 8622943
REFERENCE 4 (bases 1 to 37821)
Salazar.L., Figsli.H., de Rossi.E., Riccardi.G., Rios.C., Cole.S.T.
and Takiff.H.E.
Organization of the origins of replication of the chromosomes of
Mycobacterium smegmatis, Mycobacterium leprae and Mycobacterium

```

```

tuberculosis and isolation of a functional origin from M. smegmatis
Mol. Microbiol. 20 (2), 283-293 (1996)
96310367
8733228
5 (bases 1 to 37821)
Cole.S.T.
Direct Submission
Submitted (25-AUG-1997) Dr. Stewart T. Cole, Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France Requests for cosmids should be sent to
Karin Eigmeier (kei@pasteur.fr)
On Aug 27, 1997 this sequence version replaced gi:1262351.
Notes:
Work in Paris is supported by the Heiser Trust, the Association
Francaise Raoul Follereau and the Groupement de Recherches et des
Etudes des Genomes (GIP-GREG).
CDS are numbered using the following system eg MLCB33.01c. ML (M.
leprae), CB33 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. All CDS
over 100 codons have been analysed. Gene prediction is based on
positional base preference in codons especially where there is an
increase in the observed/expected third position G + C. CAUTION:
We may not have predicted the correct initiation codon. Where
possible we choose an initiation codon (atg, gtg, or ttg) which is
preceded by an upstream ribosome binding site sequence (optimally
5-13bp before the initiation codon). If this cannot be identified
we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid B1770
overlaps cosmids L222 and B628 at the 5' and 3'-ends, respectively.
Location/Qualifiers
1..37821
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/db_xref="taxon:1769"
/clone="cosmid B1770"
1..598
/gene="dnaA"
<1..598
/gene="dnaA"
/notes="MLB1770.01; dnaA"; The 5' part of the M. leprae
dnaA gene is on cosmid L222 (accession number L39923)"
/codon_start=2
/transl_table=11
/product="DnaA"
/protein_id="CAA94708.1"
/db_xref="GI:1262352"
/translation="QLATLEDRRLTRFPGHGLTDVQPELTRLAIRKKAKMERLAV
PDVUELASSIERNIREGALIRVTFASLNKTAIDKALAEIVLRDLIDASTMQI
SAATIMTATAYFTTIELRPGKTRALAQSRQIAMYLCLRELTDLSLPKTGQAFGRD
HTTVYAQRKILSEMAERREVDHVKELTTRIRQSKK"
1113..2312
/gene="dnaN"
1113..2312
/gene="dnaN"
/notes="MLB1770.02; dnaN; len: 399"
/codon_start=1
/transl_table=11
/product="DnaN"
/protein_id="CAA94709.1"
/db_xref="GI:1262353"
/db_xref="GOA:P46387"
/db_xref="SWISS-PROT:P46387"
/translation="MDLAKTNVGCSDLKFLCLARESFASVSWAKYLFTRPTVPVLISG

```

```
VLGTGDSGLTSGFDYVSAEYVAAEIASGSSVLVSGRLSDITRALPNKPVHYV
DGNRALTCGSARFSPTMAVEDYPTLPTDPTGTLPSDFAEAGVIAAGRDYIT
LPMLTGIRIEISGDTVLAATFRVLAVERLKWLSVSSDFEASVLVPAKTLVEVAKAG
TDGSGVCLGAGVGKDGKLGISGGKRSTRLLDABFPKFROLLPAETHVATATID
VAELTAIKLVALVADRGACQVMEFGDGLRLSAGADDDVGRAEEDLAVAFTEPTIA
FPNYLTDLGLASVHSERSVFGFTTSPKALLRPTSDNDVHPHTDGFPPALPDYVILL
MPVRLFG"
2313..3470
/gene="recF"
2313..3470
/gene="recF"
/notes="MLB1770.03, recF, len: 385"
/codon_start=1
/transl_table=11
/product="RecF"
/protein_id="CAA94710.1"
/db_xref="GI:12623354"
/db_xref="GOA:P46391"
/db_xref="SWISS-PROT:P46391"
/translations="MYVRRHGLRDFRSWDHVDLELNPGRVTFPGNGKXKNLIEALW
YSTLSHRVGTDIPIRAGTIRAIIVTVNRECAIDLEIAGRANRNLRSLV
RGNREVGLVAVLPEPDLALVCGDPANRRYLDLATVQPVTAARVADVKVLRQ
RLALKSLAARYESDQGLDITDVMWTRIAEHAELMAARIDLNVQLAPEYKAYOL
LAPGSTASTSYRASLDIGGIAGVSSDRALLQADLLAGLSTRNVEIERGICLVGPH
RDELELRGDPQAKGFASGHGSWSLAIRLAAYELLRADGNPEVLLDDVFAELDA
RCALATVAESAEOVLVTSNAQEDIPVGNDAKWTVDLRSDSGRVSVPY"
3454..3459
/gene="recF"
/notes="possible RBS, GGTGG, for MLB1770.04"
3467..4066
/gene="MLB1770.04"
3467..4066
/gene="MLB1770.04"
/notes="MLB1770.04, unknown, previously orf199, len: 199 ;
encodes hypothetical protein in the RecF-GyrB intergenic
region"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAA94711.1"
/db_xref="GI:1262335"
/db_xref="SWISS-PROT:Q9CDF4"
/translations="MIESNESYGGDTIEPLTGLSGFDLVRRALEAPAAACAQGD
AGHVGVPVFPVTRNRNWSGPDVDPQLGVKVAHDLAKRQWSQVAEGRVFG
QMASVGGQADAHAFVGLNNGVLSYTABSTAWATOLRIMOQALLAKIAAAVGVVVT
SLKTIPTAPSRKRWKPHIAGRCPRTTASRHAGSELNR"
4195..4198
/notes="possible RBS, GGAG, for gyrB"
4203..6296
/gene="gyrB"
4203..6296
/gene="gyrB"
/notes="MLB1770.05, gyrB, len: = 697"
/codon_start=1
/transl_table=11
/product="DNA gyrase B"
/protein_id="CAA94712.1"
/db_xref="GI:1262336"
/db_xref="GOA:Q59533"
/db_xref="SWISS-PROT:Q59533"
/translations="MTAAVTPGLTCLNKESIQVIAQRKQADYEYGAASITILEAV
RKPGMYVGTSGERGLHLIWEVDNSVDMAGYATQVDVRLFDGDSVEVADNGRI
PVAVHATGVTQVMTVTHAGGFGKSGVSGGLHGVSVVNALSTRVYEDIK
RDGYENSQFYDKAVPGILKQGEATEATGTRIFWADPDIFETTKYDFGVARRTHYPA
FLNKGILTINLVDREVKQDVVDVSDTAEPVAMTVEEKSTESSAPKRVHRTFYP
GGVLDVFVKNTRKTPQOSI IDFDGKGAGHEVEVAMQWNGYSESVHVFANTINPHE
GKTHGEGFRALTSVKNYAKDKLLKQDPNLITGDDIREGLAAVLSKVSPQPEQV
TKTKLGNTEVKSIVQVNCNQLHWFANFVDAVVNKAISQAQARIARAKRELVR
RKSNADIGLPGKADCRDTPRSSELYVVEGDSAGKSAGSDSMFQAILPGRKII
NVEKARIDVLYKNTQVQAIITALGTGTHDEFIDSLRYRHKVILMADADVGGHISTLL
LTLFLRMRPLIEHGYFLAQPPLYLKWQMDPEPAYSDSERDGLLETGLKLGKIN
KEDGLOQYKGLGMDAKELWETTMDFPSVRVLRQVTLDDAAAADDELFSILMGEDVDAR
SFTTRNAKDVRFIDV"
```

```
6340..6345
/notes="possible RBS, GAGGAA, for gyrA"
6349..10170
/gene="gyrA"
6349..10170
/gene="gyrA"
/notes="MLB1770.06, gyrA, encodes a GyrA precursor protein
spliced into a mature A subunit of DNA gyrase and a free
protein intron (intein); len: 1273"
/codon_start=1
/transl_table=11
/product="DNA gyrase A, and intein"
/protein_id="CAA94713.1"
/db_xref="GI:1262357"
/db_xref="GOA:Q57532"
/db_xref="SWISS-PROT:Q57532"
/translations="MTDITLPPGDSIQVSEVPDIQEQMQRSYIDYAMSVIVGRALPE
VRDLGKPVHRRVLYAMLDGFRPDRSHAKSARSAETMGVNHPHGDASITDVLVRMAQ
Query Match 79.3%; Score 23.8; DB 1; Length 37821;
Best Local Similarity 92.6%; Pred. No.12; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 2;
QY 3 GGCACAGAACCCAGGAGTAGCACCACATGA 29
|||||
Db 27225 GGCACAGAACCCAGGAGTAGCACCACATGA 27199
|||||

RESULT 6
MLEPRTN1/c
LOCUS
DEFINITION
Mycobacterium leprae strain TN complete genome, segment 1/10.
ACCESSION
AL583917 AL583917.1 GI:13092412
VERSION
AL583917.1 GI:13092412
KEYWORDS
SOURCE
Mycobacterium leprae
ORGANISM
Mycobacterium leprae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
1 (bases 1 to 344050)
Cole, S.F., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,
Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,
Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,
Hamil, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C.,
Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,
Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,
Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K.,
Taylor, K., Whitehead, S., Woodward, J.R. and Barrell, B.G.
Massive gene decay in the leprosy bacillus
Nature 409 (6823), 1007-1011 (2001)
21128732
PUBMED
11234002
2 (bases 1 to 344050)
Parkhill, J.
Direct Submission
Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
Notes:
Details of M. leprae sequencing at the Sanger Centre are available
from http://www.sanger.ac.uk/Projects/M_leprae/ A relational
database containing the M. leprae sequences is available from
http://genolist.pasteur.fr/Lepraoma/.
Location/Qualifiers
1..344050
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/strain="TN"
/db_xref="taxon:1769"
1..1566
source
gene
```

```

/gene="dnaA"
/notes="synonym: ML0001"
1..1566
/gene="dnaA"
/notes="Similar to M. tuberculosis dnaA, chromosomal
replication initiator protein, SW:DNAA_MYCTU (P49593) (507
aa); Fasta score E(): 0, 87.2% identity in 507 aa overlap
and Mycobacterium smegmatis dnaA, SW:DNAA_MYCSM (P49992)
(495 aa); Fasta score E(): 0, 69.2% identity in 441 aa
overlap. Previously sequenced as SW:DNAA_MYCLE (P46388)
(521 aa); Fasta score E(): 0, 99.8% identity in 521 aa
overlap. Contains Pfam match to entry PF00308 bac_dnaA,
Bacterial dnaA protein. Contains PS00017 ATP/GTP-binding
site motif A (P-loop). Contains PS01008 DnaA protein
signature."
/codon_start=1
/transl_table=11
/product="putative chromosomal replication initiator
protein"
/protein_id="CAC29509.1"
/db_xref="GI:13092413"
/db_xref="GOA:P46388"
/db_xref="SWISS-PROT:P46388"
/translation="MFVPHAKKPEIYENQRTDSLADLSLGTFTVMNAVSENLGESH
TDDATNDSTLVLTPOQRAWLNVQPLTIIIGFALLSVPSFVQNEIERHLRTPIIT
DALSRRLGOQLGVLPSTPHIDNSSADVLITDDCGTDTDNYGELTGEYCG
LPTVTERPHHTSTVGTGSLAPRYTFTEFVIGASNEFAHAALAEAPAYNPL
FIWSEGLKHLHAAAGNTAORLFFGMRVKNYSTETFTNDNFINSLRDRKVAFKRSY
RDVLLVDDIIFEGKEGQEEFFHTNLNANKQIVISSDRPPKQATLEDRLT
RFENLITDQVPPLETRIIALRKQMERLAVFGDVLIELIASSIERNLEALIR
VTAFAKNTAIDKALAEIVLRDLIADASTMOISAAITMTATAEYDTTIEELRGPGK
TRAAQKQIAMYLURELTDLSPKIQAGFRDHTTWYAKRILSEMAERREVFQHV
KELTRIRORSKR"
544..1482
/gene="dnaA"
/notes="Pfam match to entry PF00308 bac_dnaA, Bacterial
dnaA protein, score 712.20, E-value 8.3e-240"
664..687
/gene="dnaA"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
1426..1482
/gene="dnaA"
/notes="PS01008 DnaA protein signature"
2081..3280
/gene="dnaA"
/notes="synonym: ML0002"
2081..3280
/gene="dnaA"
/EC-numbers="2.7.7.7"
/notes="Similar to M. tuberculosis dnaA, DNA polymerase
III, [beta] subunit, SW:DP3B_MYCTU (Q50790) (402 aa);
Fasta score E(): 0, 80.9% identity in 403 aa overlap and
Mycobacterium smegmatis dnaA, SW:DP3B_MYCSM (P52851) (397
aa); Fasta score E(): 0, 77.3% identity in 397 aa overlap.
Previously sequenced as SW:DP3B_MYCLE (P46387) (399 aa);
Fasta score E(): 0, 100.0% identity in 399 aa overlap.
Contains Pfam match to entry PF00712 DNA_pol3_beta, DNA
polymerase III beta subunit."
/codon_start=1
/transl_table=11
/product="putative DNA polymerase III, [beta] subunit"
/protein_id="CAC29510.1"
/db_xref="GI:13092414"
/db_xref="GOA:P46387"
/db_xref="SWISS-PROT:P46387"
/translation="MDIAKTVNGSDLKFKCLARESFASVSWAKYLPTRPTVPVLSSG
VLTGSGSLTISGFVSAEVAIASSGVLVSGRLSDITRALPKPVPVY
LPMGLVTCGARSFSLPTMAVEDYPTLPTLPDRTGTLPSDVFAEAGVIAAGRDYV
DGNELTARISGSDTIVLAATDPRLAVERLKNVSLSSDPEASVLPVAKTJNEVAFG
TDGSGVCLISGVNGKGLGFLISGGKRSSTRLLDAPFPKPKQLLPHEHTAVATID
VAETEAIKVALVADRGAVQRMFEGDILRLSAGADVGRAEEDLAVAPTGEPLTIA
FNNPLYLTDGLASVHSERSVFGFTTSPKALLRPTSDNDVHTHDGFPALPTDYVLL
MPVRLPG"

```

```

2117..3214
/gene="dnaA"
/notes="Pfam match to entry PF00712 DNA_pol3_beta, DNA
polymerase III beta subunit, score 592.70, E-value
2.2e-174"
3281..4438
/gene="recP"
/notes="synonym: ML0003"
3281..4438
/gene="recP"
/notes="Similar to M. tuberculosis recP, DNA replication
and SOS induction protein, SW:RECF_MYCTU (Q59586) (385
aa); Fasta score E(): 0, 76.4% identity in 385 aa overlap
and Mycobacterium smegmatis recP, SW:RECF_MYCSM (P50916)
(384 aa); Fasta score E(): 0, 70.0% identity in 383 aa
overlap. Previously sequenced as SW:RECF_MYCLE (P46391)
(385 aa); Fasta score E(): 0, 99.5% identity in 385 aa
overlap. Contains 2 Pfam matches to entry PF00470 RecF,
RecF protein. Contains PS00617 RecF protein signature 1.
Contains PS00017 ATP/GTP-binding site motif A (P-loop).
Contains PS00618 RecF protein signature 2."
/codon_start=1
/transl_table=11
/product="putative DNA replication and SOS induction
protein"
/protein_id="CAC29511.1"
/db_xref="GI:13092415"
/db_xref="GOA:P46391"
/db_xref="SWISS-PROT:P46391"
/translation="MYVHFGFLRFRSDVHDLNLPRTVFFGPNNGKNTLIEALW
XSTLSSHRVGTDLIRAGTIRAIIVTVNREGCAIDLEIAAGRANRANRSLV
RGMRVGVGLRAVLPAPEDLALVCGDDPANRRYLLDIAVTQPVYADYDKVLRQ
RTALKLSAAARYSDQGVQDLTDVWDTRLAEHGAELMAARI DLVNLGLAPEVEAYOL
LAPCSRTASISYRASLDIGGIAGYSSDRALLQDLLAGI STRENVEISGICLVGPH
RDEELRUGDQPAKFGAFSGHSGSLATRLAAYELLRADNCFVLLDDVFAELDAA
RCRLATVAESAQVLTSAAGDIPVGWDARVTVLDLSDSGRSVSVYP"
3287..3415
/gene="recP"
/notes="Pfam match to entry PF00470 RecF, RecF protein,
score 18.90, E-value 0.00024"
3368..3391
/gene="recP"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
3614..3691
/gene="recP"
/notes="PS00617 RecF protein signature 1"
4100..4423
/gene="recP"
/notes="Pfam match to entry PF00470 RecF, RecF protein,
score 5.20, E-value 1.5"
4244..4297
/gene="recP"
/notes="PS00618 RecF protein signature 2"
4435..5004
/gene="ML0004"
/notes="ML0004"
/notes="Similar to hypothetical proteins from mycobacteria
e.g. M. tuberculosis RV0004, hypothetical protein,
TR:P71573 (EMBL:AL123456) (187 aa); Fasta score E(): 0,
77.2% identity in 167 aa overlap. Previously sequenced as
TR:Q50181 (EMBL:Z70722) (199 aa); Fasta score E(): 0,
99.5% identity in 187 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC29512.1"
/db_xref="GI:13092416"
/db_xref="SWISS-PROT:Q9CDF4"
/translation="MIESNESYSYGGDTIIEPLGTLSGFDLVRALREARAAACGQKD
AGTVHVPFPVFRVTRRRNWSGPDVDRPQPLGKVAHDLAKKGRSAQVAERGVFG
QWASNVGQIADHAPFVGLNNNGVLSVTAETAWATQIRIMQALLAKIAAAGVGVIT
SLKITGTPASWRKGPWHIAGRPRDTYG"

```


shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP11-561G1 is at 2000 in this sequence.

----- Summary Statistics -----
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315

FEATURES
source
1. 96512
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-456A18"
/clone_lib="RP11-11"

ORIGIN
Query Match 70.7%; Score 21.2; DB 9; Length 96512;
Best local similarity 88.5%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGCGAGAACCGAGGAGTAGACCAA 26
DB 62133 TAGGCGAGAACCGAGGAGTAGACCAA 62108

RESULT 8
AC021718 154359 bp DNA linear HTG 12-MAR-2000
LOCUS Homo sapiens chromosome 11 clone RP11-207E21 map 11, WORKING DRAFT
DEFINITION
SEQUENCE 15 unordered pieces.
AC021718 3 GI:7229845
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154359)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-207E21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154359)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, P., Boguslavsky, L., Boukhtalter, B., Brown, A., Burkett, G., Castle, A., Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lebecky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meidrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6970561.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

5211..5216
/note="possible RBS"
5229..7265
/gene="gyrB"
/note="synonym: ML0005"
5229..7265
/gene="gyrB"
/EC numbers="5.99.1.3"
/note="Similar to M. tuberculosis gyrB, DNA gyrase subunit B, SW:GYRB MYCTU (P41514) (686 aa); Pasta score E(): 0, 87.5% identity in 679 aa overlap and to Mycobacterium

Query Match 79.3%; Score 23.8; DB 1; Length 344050;
Best local similarity 92.6%; Pred. No. 14;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCGCAGAACCGAGGAGTAGACCAATGA 29
DB 28197 GCGCAGAACCGAGGAGTAGACCAATGA 28171

RESULT 7
AC022539/c 96512 bp DNA linear PRI 09-APR-2002
LOCUS Homo sapiens chromosome 10 clone RP11-456A18, complete sequence.
DEFINITION
AC022539
ACCESSION AC022539
VERSION AC022539.13 GI:20087101
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 96512)
AUTHORS Smith, D. R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Sequence Data
Unpublished
REFERENCE 2 (bases 1 to 96512)
AUTHORS Smith, D. R.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 96512)
AUTHORS Smith, D. R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 96512)
AUTHORS Smith, D. R.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 5 (bases 1 to 96512)
AUTHORS Smith, D. R.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 6 (bases 1 to 96512)
AUTHORS Smith, D. R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
On Apr 9, 2002 this sequence version replaced gi:19774371.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg142

IMPORTANT: This sequence is not the entire insert of clone RP11-456A18. It may be

Contact: sequence_submission@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5944
 Center clone name: 207_E_21
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 147423 bases at least Q40
 Consensus quality: 150978 bases at least Q30
 Consensus quality: 152145 bases at least Q20
 Insert size: 158000; agarose-1p
 Quality coverage: 4.3 in Q20 bases; agarose-1p
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1155: contig of 1155 bp in length
 * 1156 1255: gap of 100 bp
 * 1256 3249: contig of 1994 bp in length
 * 3250 3349: gap of 100 bp
 * 3350 5935: contig of 2586 bp in length
 * 5936 6035: gap of 100 bp
 * 6036 8676: contig of 2641 bp in length
 * 8677 8777: gap of 100 bp
 * 8778 11533: contig of 2757 bp in length
 * 11534 11634: gap of 100 bp
 * 11635 15331: contig of 3698 bp in length
 * 15332 15431: gap of 100 bp
 * 15432 19618: contig of 4187 bp in length
 * 19619 19718: gap of 100 bp
 * 19719 26030: contig of 6312 bp in length
 * 26031 26130: gap of 100 bp
 * 26131 34060: contig of 7930 bp in length
 * 34061 34160: gap of 100 bp
 * 34161 43748: contig of 9588 bp in length
 * 43749 43849: gap of 100 bp
 * 43849 55829: contig of 11981 bp in length
 * 55830 55930: gap of 100 bp
 * 55931 68307: contig of 12378 bp in length
 * 68308 68407: gap of 100 bp
 * 68408 82534: contig of 14127 bp in length
 * 82535 82635: gap of 100 bp
 * 82635 102989: contig of 20354 bp in length
 * 102989 103089: gap of 100 bp
 * 103089 154359: contig of 51271 bp in length.
 FEATURES
 source
 1..154359
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-207E21"
 /clone_lib="RPCI-11 Human Male BAC"
 1..1155
 /note="assembly_fragment"
 clone_end:17
 vector_side:left
 1256..3249
 /note="assembly_fragment"
 3350..5935
 /note="assembly_fragment"
 6036..8676
 /note="assembly_fragment"
 8777..11533
 /note="assembly_fragment"
 misc_feature
 1256..3249
 /note="assembly_fragment"
 misc_feature
 3350..5935
 /note="assembly_fragment"
 misc_feature
 6036..8676
 /note="assembly_fragment"
 misc_feature
 8777..11533
 /note="assembly_fragment"

misc_feature
 11634..15331
 /note="assembly_fragment"
 misc_feature
 15432..19618
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 19719..26030
 /note="assembly_fragment"
 26131..34060
 /note="assembly_fragment"
 34161..43748
 /note="assembly_fragment"
 43849..55829
 /note="assembly_fragment"
 55930..68307
 /note="assembly_fragment"
 68408..82534
 /note="assembly_fragment"
 82635..102988
 /note="assembly_fragment"
 103089..154359
 /note="assembly_fragment"
 ORIGIN
 Query Match 70.7%; Score 21.2; DB 2; Length 154359;
 Best Local Similarity 88.3%; Pred. NO. 2e+02; 3; Indels 0; Gaps 0;
 Matches 23; Conservative 0; Mismatches 3;
 QY 1 TGGCGCAGAACCCAGGAGTAGACCAA 26
 Db 151077 TAGGCGCAGAACCCAGGAGTAGACCAA 151102
 RESULT 9
 AC091772 158559 bp DNA linear PRI 17-JUN-2003
 LOCUS Homo sapiens chromosome 10 clone RP11-257L6, complete sequence.
 DEFINITION AC091772
 ACCESSION AC091772
 VERSION AC091772.5 GI:31795104
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 158559)
 AUTHORS Smith,D.R.
 TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 158559)
 AUTHORS Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-2001) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02453, USA
 REFERENCE 3 (bases 1 to 158559)
 AUTHORS Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (06-NOV-2001) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02453, USA
 REFERENCE 4 (bases 1 to 158559)
 AUTHORS Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (07-APR-2003) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02453, USA
 REMARK Sequence updated to remove bacterial transposon
 REFERENCE 5 (bases 1 to 158559)
 AUTHORS Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUN-2003) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02453, USA
 REMARK Sequence updated to remove vector contamination
 COMMENT On Jun 17, 2003 this sequence version replaced gi:29570805.

```

FEATURES
  source
    Location/Qualifiers
      1..158559
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="10"
        /clone="RP11-257L6"
        /clone_lib="RPC1-11"
ORIGIN
  Query Match
  Best Local Similarity 70.7%; Score 21.2; DB 9; Length 158559;
  Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TGGGCGAGAACCCAGGAGTAGACCA 26
Db 132869 TAGGCGAGAACCCAGGAGTAGACCA 132894

RESULT 10
AC069541/c
LOCUS
  DEFINITION Homo sapiens chromosome 10 clone RP11-38902, complete sequence.
  ACCESSION AC069541
  VERSION AC069541.5 GI:20330781
  KEYWORDS HTG.
  SOURCE Homo sapiens (human)
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 158568)
    Smith,D.R.
    Genome Therapeutics Corporation Sequencing Center: Human Genome
    Sequence Data
  JOURNAL
    Unpublished
  REFERENCE
    2 (bases 1 to 158568)
    Smith,D.R.
    Direct Submission
  JOURNAL
    Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
    Street, Waltham, MA 02453, USA
  REFERENCE
    3 (bases 1 to 158568)
    Smith,D.R.
    Direct Submission
  JOURNAL
    Submitted (13-APR-2001) Genome Therapeutics Corporation, 100 Beaver
    Street, Waltham, MA 02453, USA
  REFERENCE
    4 (bases 1 to 158568)
    Smith,D.R.
    Direct Submission
  JOURNAL
    Submitted (26-APR-2002) Genome Therapeutics Corporation, 100 Beaver
    Street, Waltham, MA 02453, USA
  REMARK
    1329 bases of Tn10 (J01829) transposon removed here
  COMMENT
    On Apr 26, 2002 this sequence version replaced gi:13621215.
  FEATURES
    source
      Location/Qualifiers
        1..158568
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /chromosome="10"
          /clone="RP11-38902"
          /clone_lib="RPC1-11"
ORIGIN
  Query Match
  Best Local Similarity 70.7%; Score 21.2; DB 9; Length 158568;
  Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TGGGCGAGAACCCAGGAGTAGACCA 26
Db 25700 TAGGCGAGAACCCAGGAGTAGACCA 25675

RESULT 11
AC118678/c
LOCUS
  DEFINITION Mus musculus clone RP23-406A5, WORKING DRAFT SEQUENCE, 11 unordered
  pieces.
  ACCESSION AC118678
  VERSION AC118678.3 GI:29164651
  KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
  SOURCE Mus musculus (house mouse)
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE
    1 (bases 1 to 188591)
    Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
    Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
    Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
    Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
    Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
    Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
    Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
    Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
    Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
    Landers,T., Lehorzky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
    MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
    McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
    Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
    Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
    Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
    Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
    Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
    Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
    Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
    Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
    Direct Submission
  JOURNAL
    Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
  REFERENCE
    3 (bases 1 to 188591)
    Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
    Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
    Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
    Collymore,A., Cook,A., Cooke,P., Corum,B., DeAtellano,J.,
    Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
    Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
    Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
    Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
    Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
    Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
    Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
    Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
    Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
    O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
    Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
    Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
    Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
    Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
    Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
    Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
    Direct Submission
  JOURNAL
    Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT
    On Mar 23, 2003 this sequence version replaced gi:28269643.
    All repeats were identified using RepeatMasker:
    Smit,A.P.A. & Green,P. (1996-1997)
    http://ftp.genome.washington.edu/RN/RepeatMasker.html
    ----- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu

```

```

----- Project Information
Center project name: L19876
Center clone name: 406_A_5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186196 bases at least Q40
Consensus quality: 186891 bases at least Q30
Consensus quality: 187169 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 187591; sum-of-contigs
Quality coverage: 9.7 in Q20 bases; agarose-fp
Quality coverage: 9.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5777: contig of 5777 bp in length
* 5778 5877: gap of 100 bp
* 5878 7818: contig of 1941 bp in length
* 7818 7918: gap of 100 bp
* 7918 11201: contig of 3283 bp in length
* 11201 11301: gap of 100 bp
* 11301 11882: contig of 7581 bp in length
* 11882 18983: gap of 100 bp
* 18983 27416: contig of 8434 bp in length
* 27416 27516: gap of 100 bp
* 27516 48440: contig of 21024 bp in length
* 48440 48541: gap of 100 bp
* 48541 65525: contig of 17885 bp in length
* 65525 66526: gap of 100 bp
* 66526 93177: contig of 26552 bp in length
* 93177 93278: gap of 100 bp
* 93278 114525: contig of 21348 bp in length
* 114525 114725: gap of 100 bp
* 114725 143496: contig of 28770 bp in length
* 143496 143595: gap of 100 bp
* 143595 188591: contig of 44396 bp in length.
*
* Location/Qualifiers
*
* 1. 188591
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP23-406A5"
* /clone_lib="RPCI-23 Female Mouse BAC"
*
* 1. 5777
* /note="assembly_fragment"
* clone_end:896
* vector_side:left"
*
* misc_feature
* 5878..7818
* /note="assembly_fragment"
* 7919..11201
* /note="assembly_fragment"
* 11302..18882
* /note="assembly_fragment"
* 18983..27416
* /note="assembly_fragment"
* 27517..48540
* /note="assembly_fragment"
* 48641..66525
* /note="assembly_fragment"
* 66626..93177
* /note="assembly_fragment"
* 93278..114525
* /note="assembly_fragment"
* 114726..143495
* /note="assembly_fragment"

```

```

misc_feature 143596..188591
/note="assembly_fragment"
clone_end:17
vector_side:right"

ORIGIN

Query Match 70.0%; Score 21; DB 2; Length 188591;
Best Local Similarity 82.8%; Pred. No. 2.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 30
|||||
Db 135354 GGCAAAACCCAGGAGGACGACCAATGAG 135326

RESULT 12
AC105077/c
LOCUS
DEFINITION
Mus musculus clone RP23-390G1, WORKING DRAFT SEQUENCE, 8 ordered
pieces
AC105077
VERSION
AC105077.4 GI:22381833
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 208342)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-390G1
Unpublished
REFERENCE
2 (bases 1 to 208342)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Grandpierre,N.,
Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208342)
REFERENCE
AUTHORS
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepl,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Grandpierre,N., Grand-Pierre,N., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,

```



```

PC C12N5/10.C12P21/02.C12P21/08.C12Q1/68.C12N15/00.C12N5/00, PC
C12N15/00
CC Von Heijne matrix
CC score 7.6999980926514
CC seq LLFLLALTGLLL/LV
FH Key Location/Qualifiers
FT CDS 58..363
FT sig_peptide 58..111.
FT Location/Qualifiers
FEATURES
    source
        1..510
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 510;
Best Local Similarity 85.2%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 125 TGGCGCTGACCCAGGAGTAGCAGAGT 99

RESULT 15
E10616/c
LOCUS Human cDNA encoding cytochrome P4502B6.
DEFINITION
E10616
ACCESSION
E10616
VERSION
E10616.1 GI:22027672
KEYWORDS
JP 1996027193-A/3.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1476)
AUTHORS
Funae,Y., Imaoka,S., Matsuki,Y., Hayashi,K. and Yabusaki,Y.
TITLE
ANTIBODY RECOGNIZING CYTOCHROME P4502B6 ORIGINATED FROM MAN
JOURNAL
SUMITOMO CHEM CO LTD
COMMENT
OS Homo sapiens (human)
PN JP 1996027193-A/3
PD 30-JAN-1996
PF 13-JUL-1994 JP 1994161548
PI FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
YABUSAKI YOSHIYASU
PC C07K16/18,C12N15/09//C12N1/19,C12N9/02,G01N33/53,(C12N1/19, PC
C12R1:865);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1476
FT FT /organism='Homo sapiens'
FT FT /tissue type='liver'
FT CDS 1..1476
FT FT /product='cytochrome P4502B6'.
FT Location/Qualifiers
FEATURES
    source
        1..1476
        /organism="Homo sapiens"
        /mol_type="genomic RNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 68 TGGCGCTGACCCAGGAGTAGCAGAGT 42

PC C12N5/10.C12P21/02.C12P21/08.C12Q1/68.C12N15/00.C12N5/00, PC
C12N15/00
CC Von Heijne matrix
CC score 7.6999980926514
CC seq LLFLLALTGLLL/LV
FH Key Location/Qualifiers
FT CDS 58..363
FT sig_peptide 58..111.
FT Location/Qualifiers
FEATURES
    source
        1..510
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 510;
Best Local Similarity 85.2%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 125 TGGCGCTGACCCAGGAGTAGCAGAGT 99

RESULT 16
E10861/c
LOCUS cDNA encoding human cytochrome P450.
DEFINITION
E10861
ACCESSION
E10861
VERSION
E10861.1 GI:22027955
KEYWORDS
JP 1996056695-A/10.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1476)
AUTHORS
Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
Nakatsuka,I.
TITLE
METHOD FOR EVALUATING SAFETY
JOURNAL
SUMITOMO CHEM CO LTD
COMMENT
OS Homo sapiens (human)
PN JP 1996056695-A/10
PD 05-MAR-1996
PF 15-JUL-1994 JP 1994164184
PR 20-JUL-1993 JP 93P 201120, 30-JUL-1993 JP 93P 208279, PR
17-JUN-1994 JP 94P 136053
PI HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI KOMAI
KOICHIRO,
KI KANEKO HIDEO, NAKATSUKA IWA0
PC C12Q1/02.C12M1/34,C12Q1/26;
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FT source 1..1476
FT FT /organism='Homo sapiens'
FT CDS 1..1476
FT FT /product='human cytochrome P450 2B6'.
FT Location/Qualifiers
FEATURES
    source
        1..1476
        /organism="Homo sapiens"
        /mol_type="genomic RNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 68 TGGCGCTGACCCAGGAGTAGCAGAGT 42

RESULT 17
AR399329/c
LOCUS AR399329 1476 bp DNA linear PAT 18-DEC-2003
DEFINITION
AR399329
ACCESSION
AR399329
VERSION
AR399329.1 GI:40141131
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 1476)
AUTHORS
Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
Nakatsuka,I.
TITLE
Method for safety evaluation of chemical compound using recombinant
yeast expressing human cytochrome P450
JOURNAL
Patent: US 6620593-A 10 16-SEP-2003;
FEATURES
    source
        1..1476
        /organism="unknown"
        /mol_type="genomic DNA"

```

```

ORIGIN
Query Match      68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCACCACAT 27
    |||||
Db 68 TGGCGCTGAACGAGGAGTAGCAGAGT 42

RESULT 18
AF182277/c
LOCUS      1520 bp mRNA linear PRI 29-NOV-1999
DEFINITION Homo sapiens cytochrome P450-2B6 (CYP2B6) mRNA, complete cds.
ACCESSION AF182277
VERSION AF182277.1 GI:6470142
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 1520)
AUTHORS   Wu,J., Dong,H., Yu,Y. and Zhu,L.
TITLE     Cloning and identification of human cytochrome CYP2B7 gene cDNA and
          construction of its mammalian cell expression recombinant
JOURNAL   Teratog., Carcinog. Mutagen. 7 (1), 1-6 (1995)
REFERENCE  2 (bases 1 to 1520)
AUTHORS   Zhuge,J., Qian,Y., Xie,H. and Yu,Y.
TITLE     Direct Submission
JOURNAL   Submitted (01-SEP-1999) Pathophysiology, Zhejiang University, School
          of Medicine, 353 Yan'an Road, Hangzhou, Zhejiang 310031, China
FEATURES   Location/Qualifiers
            source
              1..1520
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="19"
                /map="19q13.2"
                /tissue type="liver"
            gene
              1..1520
                /gene="CYP2B6"
            CDS
              3..1478
                /gene="CYP2B6"
                /SC_number="1.14.14.1"
                /codon_start=1
                /product="cytochrome P450-2B6"
                /protein_id="AAF13602.1"
                /db_xref="GI:6470143"
                /translation="WELSVLLFLALLTGLLLLVQRHPTNTHDLPPGPRPLPLGLNLL
                QMDRGLKSFPRFKYGVDFVTHLGPVPVLMCGVEAREALVDXAEAFSGRGKIA
                WDPFPRGVIFANGNKKVLRFSVTIMDDFGMGKRSVEERTQEEAQCILBELRKS
                KGLMDATFISFANTICISVFGKREHYQDFGLKMLNFYQTFLSIYSVFGQFPE
                LFSGLKFPQHLNRQVYNPQINAIYGHSEKREHRLDPSAPRDLIDTYLLHMEKEK
                SNAHSEFQNNLNTLGLTFAGTETTTTLRYGFLMLKYPHVAERVRYEIEOVIGP
                HRPBLHDAKMPYEAVIETQRFSDLLPMGVPHIVTQHTSPRGYIIPKDTVEFLIL
                STALHDFFKPEFADPNFDLDAKALKTEARIPPSLGRKICLGLGIARAEILFFP
                TITLQNFWSASVAPEDIDLTPQECGVKIPFTYQIRFLPR"

ORIGIN
Query Match      68.7%; Score 20.6; DB 9; Length 1520;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCACCACAT 27
    |||||
Db 70 TGGCGCTGAACGAGGAGTAGCAGAGT 44

RESULT 19
AF081569/c
LOCUS      2000 bp DNA linear PRI 20-APR-1999
DEFINITION Homo sapiens cytochrome P450 (CYP2B6) gene, partial cds.
ACCESSION AF081569
VERSION AF081569.1 GI:4587968
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 2000)
AUTHORS   Sueyoshi,T., Kawamoto,T., Zelko,I., Honkakoski,P. and Negishi,M.
TITLE     The repressed nuclear receptor CAR responds to phenobarbital in
          activating the human CYP2B6 gene
JOURNAL   J. Biol. Chem. 274 (10), 6043-6046 (1999)
MEDLINE   PUBMED 10037683
REFERENCE  2 (bases 1 to 2000)
AUTHORS   Kawamoto,T. and Negishi,M.
TITLE     Direct Submission
JOURNAL   Submitted (02-AUG-1998) Laboratory of Reproductive and
          Developmental Toxicology, National Institute of Environmental
          Health Sciences, P. O. Box 12233, Research Triangle Park, NC 27709,
          USA
FEATURES   Location/Qualifiers
            source
              1..2000
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                <1843..>2000
                /gene="CYP2B6"
                <1843..>2000
                /gene="CYP2B6"
                /product="cytochrome P450"
                1843..>2000
                /gene="CYP2B6"
                /note="subfamily IIB; phenobarbital-inducible; polypeptide
                6"
                /codon_start=1
                /product="cytochrome P450"
                /protein_id="AAD25924.1"
                /db_xref="GI:4587969"
                /translation="WELSVLLFLALLTGLLLLVQRHPTNTHDLPPGPRPLPLGLNLL
                QMDRRGLL"

ORIGIN
Query Match      68.7%; Score 20.6; DB 9; Length 2000;
Best Local Similarity 85.2%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCACCACAT 27
    |||||
Db 1910 TGGCGCTGAACGAGGAGTAGCAGAGT 1884

RESULT 20
AX018551/c
LOCUS      3045 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 45 from Patent WO9945127.
ACCESSION AX018551
VERSION AX018551.1 GI:10042692
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS   Kingsman,S.M., Mitrophanous,K., Patterson,A.V., Stratford,I.J.,
          Griffiths,L. and Kan,O.
TITLE     Enhanced prodrug activation
          Patent: WO 9945127-A 45 10-SEP-1999;
          KINGSMAN SUSAN MARY (GB); MITROPHANOUS KYRIACOS (GB); PATTERSON

```

```

ORIGIN
Query Match      68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCACCACAT 27
    |||||
Db 68 TGGCGCTGAACGAGGAGTAGCAGAGT 42

RESULT 18
AF182277/c
LOCUS      1520 bp mRNA linear PRI 29-NOV-1999
DEFINITION Homo sapiens cytochrome P450-2B6 (CYP2B6) mRNA, complete cds.
ACCESSION AF182277
VERSION AF182277.1 GI:6470142
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1520)
AUTHORS Wu,J., Dong,H., Yu,Y. and Zhu,L.
TITLE Cloning and identification of human cytochrome CYP2B7 gene cDNA and
        construction of its mammalian cell expression recombinant
JOURNAL Teratog., Carcinog. Mutagen. 7 (1), 1-6 (1995)
REFERENCE 2 (bases 1 to 1520)
AUTHORS Zhuge,J., Qian,Y., Xie,H. and Yu,Y.
TITLE Sequence of a new human cytochrome P450-2B6 cDNA
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1520)
AUTHORS Zhuge,J., Qian,Y., Xie,H. and Yu,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Pathophysiology, Zhejiang University, School
        of Medicine, 353 Yan'an Road, Hangzhou, Zhejiang 310031, China

FEATURES
source
    1..1520
        Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="19"
            /map="19q13.2"
            /tissue type="liver"
    1..1520
        /gene="CYP2B6"
    3..1478
        /gene="CYP2B6"
    /SC_number="1.14.14.1"
    /codon_start=1
    /product="cytochrome P450-2B6"
    /protein_id="AAF13602.1"
    /db_xref="GI:6470143"
    /translation="WELSVLLFLALLTGILLLVQRHNTHTDRLPPGRLPLLLGNLL
        QMDRGLKSFPRFKYGVDFVTHLGRPVVLMCGVEAREALVDVAFSGRGKIA
        WDPFPRGVIFANGNKKVLRFSVTIMDDFGMGKRSVEERTQEEAQCILBELRKS
        KGLMDATFISITANCISIVFGKRFHQDQFGLKMLNLFYQTFLSIVFGQFPE
        LFSGLKFPQHLRVYKNPQINAIYIGSHVEKREITLDPSPADRLDITVLLHMEKEK
        SNAHSEFQNNLNTLGLTFAGTETTTTLRYGFLMLKYPHVAERVRYEIEOVIGP
        HRPBLHDAKMPYEAVIETQRFSDLLPMGVPHIVTQHTSPRGYIIPKDTVEFLIL
        STALHDQFPEKPDANFDLNDGALKATKEATIPPSLGRKICLGLGIARAEILFF
        TITLQNFWSASVAPEDIDLTPQECGVGKIPFTYQIRFLPR"

ORIGIN
Query Match      68.7%; Score 20.6; DB 9; Length 1520;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCACCACAT 27
    |||||
Db 70 TGGCGCTGAACGAGGAGTAGCAGAGT 44

RESULT 19
AF081569/c
LOCUS      2000 bp DNA linear PRI 20-APR-1999
DEFINITION Homo sapiens cytochrome P450 (CYP2B6) gene, partial cds.
ACCESSION AF081569
VERSION AF081569.1 GI:4587968
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Sueyoshi,T., Kawamoto,T., Zelko,I., Honkakoski,P. and Negishi,M.
TITLE The repressed nuclear receptor CAR responds to phenobarbital in
        activating the human CYP2B6 gene
JOURNAL J. Biol. Chem. 274 (10), 6043-6046 (1999)
MEDLINE 99156873
PUBMED 10037683
REFERENCE 2 (bases 1 to 2000)
AUTHORS Kawamoto,T. and Negishi,M.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-1998) Laboratory of Reproductive and
        Developmental Toxicology, National Institute of Environmental
        Health Sciences, P. O. Box 12233, Research Triangle Park, NC 27709,
        USA

FEATURES
source
    1..2000
        Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            <1843..>2000
            /gene="CYP2B6"
            <1843..>2000
            /gene="CYP2B6"
            /product="cytochrome P450"
            1843..>2000
            /gene="CYP2B6"
            /note="subfamily IIB; phenobarbital-inducible; polypeptide
                6"
            /codon_start=1
            /product="cytochrome P450"
            /protein_id="AAD25924.1"
            /db_xref="GI:4587969"
            /translation="WELSVLLFLALLTGILLLVQRHNTHTDRLPPGRLPLLLGNLL
                QMDRRGLL"

ORIGIN
Query Match      68.7%; Score 20.6; DB 9; Length 2000;
Best Local Similarity 85.2%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCACCACAT 27
    |||||
Db 1910 TGGCGCTGAACGAGGAGTAGCAGAGT 1884

RESULT 20
AX018551/c
LOCUS      3045 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 45 from Patent WO9945127.
ACCESSION AX018551
VERSION AX018551.1 GI:10042692
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kingsman,S.M., Mitrophanous,K., Patterson,A.V., Stratford,I.J.,
        Griffiths,L. and Kan,O.
TITLE Enhanced produg activation
        Patent: WO 9945127-A 45 10-SEP-1999;
        KINGSMAN SUSAN MARY (GB); MITROPHANOUS KYRIACOS (GB); PATTERSON

```

ADAM VORN (GB); STRATFORD IAN JAMES (GB); GRIFFITHS LEIGH (GB); KAN
ON (GB); OXFORD BIOMEDICA LTD (GB)

FEATURES

source
Location/Qualifiers
1..3045
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48
|||||

RESULT 21

AX018627/c
LOCUS 3045 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 45 from Patent WO9945126.
ACCESSION AX018627
VERSION AX018627.1 GI:10042755
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

REFERENCE
AUTHORS Kingsman, S.M., Mitrophanous, K., Patterson, A.V., Stratford, I.J.,
Griffiths, L. and Kan, O.
TITLE Enhanced prodrug activation
JOURNAL Patent: WO 9945126-A 45 10-SEP-1999;

KINGSMAN SUSAN MARY (GB); MITROPHANOUS KYRIACOS (GB); PATTERSON
ADAM VORN (GB); STRATFORD IAN JAMES (GB); GRIFFITHS LEIGH (GB); KAN
ON (GB); OXFORD BIOMEDICA LTD (GB)

FEATURES

source
Location/Qualifiers
1..3045
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48
|||||

RESULT 22

AX332736/c
LOCUS 3045 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3245 from Patent WO0194629.
ACCESSION AX332736
VERSION AX332736.1 GI:18123370
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
JOURNAL Patent: WO 0194629-A 3245 13-DEC-2001;

Avalon Pharmaceuticals (US)
Location/Qualifiers
1..3045

FEATURES

source

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48
|||||

RESULT 23

AX409672/c
LOCUS 3045 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2319 from Patent WO0229103.
ACCESSION AX409672
VERSION AX409672.1 GI:21442377
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

REFERENCE
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2319 11-APR-2002;
GENE LOGIC INC (US)

FEATURES
Location/Qualifiers
1..3045
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="EMBL/GenBank Accession No. M29874"

ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48
|||||

RESULT 24

BD136285/c
LOCUS 3045 bp DNA linear PAT 18-SEP-2002
DEFINITION Enhancement of prodrug activation.
ACCESSION BD136285
VERSION BD136285.1 GI:23231230
KEYWORDS JP 2002505341-A/38.
SOURCE Homo sapiens (human)
ORGANISM

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3045)
Stratford, I.J., Patterson, A.V., Kingsman, S.M., Kan, O., Griffiths, L.
and Mitrophanous, K.

TITLE Enhancement of prodrug activation
JOURNAL Patent: JP 2002505341-A 38 19-FEB-2002;
OXFORD BIOMEDICA LTD
COMMENT OS Homo sapiens (human)
PN JP 2002505341-A/38
PD 19-FEB-2002

PF 05-MAR-1999 JP 2000534557
PR 06-MAR-1998 GB 9804841.6, 19-AUG-1998 GB 9818103.5 PR
29-JAN-1999 GB 9902081.0
PI IAN JAMES STRATFORD, ADAM VORN PATTERSON, SUSAN MARY KINGSMAN, ON
PI KAN,

PI LEIGH GRIFFITHS, KYRIACOS MITROPHANOUS
PC A61K47/48,A61K35/76,A61K38/44,A61K45/00,A61K48/00,A61P9/10, PC
A61P29/00,
PC A61P35/00,A61P43/00,C12N5/10,C12N7/00,C12N9/02,C12N15/09, PC
A61K37/50,
PC C12N5/00,C12N15/00
CC Enhancement of prodrug activation
FH Key Location/Qualifiers
FT source 1..3045 /organism="Homo sapiens (human)".
FT Location/Qualifiers
1..3045
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48
|||||

RESULT 25
HUMCYP2BB/c 3045 bp mRNA linear PRI 02-NOV-1994
LOCUS Human cytochrome P450-IIB (h11B1) mRNA, complete cds.
DEFINITION M29874 J02864
ACCESSION M29874.1 GI:181295
VERSION cytochrome P450; cytochrome P450 IIB.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3045)
AUTHORS Yanano,S., Nambu,P.T., Aoyama,T., Meyer,U.A., Inaba,T.,
Kaiow,W., Gelboin,H.V., McBride,O.W. and Gonzalez,F.J.
TITLE cDNA cloning and sequence and cDNA-directed expression of human
P450 IIB1: identification of a normal and two variant cDNAs derived
from the CYP2B locus on chromosome 19 and differential expression
of the IIB mRNAs in human liver
JOURNAL Biochemistry 28 (18), 7340-7348 (1989)
MEDLINE 90057429
PUBMED 2573390
COMMENT Original source text: Human liver, cDNA to mRNA.
FEATURES
source
Location/Qualifiers
1..3045
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="19q13.2"
1..3045
/gene="CYP2B"
/name="CYP2B"
7..1482
/note="cytochrome P450-IIB"
/codon_start=1
/protein_id="AA52144.1"
/db_xref="GI:181296"
/db_xref="GDB:G00-120-752"
/translation="MELSVLLFLALLTGLLLLVQRHNPHTDLPPGPRPLPLGNLL
QMDRGLKFLPRFKYGVFTVHLGPRVVMCGVEAIRALVDKAEAFSGRGRKIA
KVDPFFRGYVIFANGNRKVLRRFSVTTWRDFMGKRSVEERIOBEACCLIELRKS
KGLMDPTFLIICISIVFGKRFYQDQDFLKLMLNFYQTFSLSSVFGQLFE
LFSGLFYFQANROVYNLOEINAYIGHSVKRETLDPSPAKDLIDTVLLHMEKEK
SNAHSEFQHNALNTLSIFRAGTETSTTLRYGELLMLKYPHVARVREIVGTP
HRPELHDKRMKPYTEAVIYEQRPDLLEMGVPHIVTQTSRGTIIIPKDETVFLIL
STALDHPHYEKFDPFNPFDLNGALKTEAFIPFSLGRICLGEIGARAEFLFF
TTLQNFSPMASPVAPEDIDLTPOEGVGKIPPTYQIRFLPR"
1716..1997
repeat_region

repeat_region 2134..2306
/note="Alu repeat 1"
ORIGIN
Query Match 68.7%; Score 20.6; DB 9; Length 3045;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48
|||||

RESULT 26
AC011541/c 39944 bp DNA linear PRI 20-JUL-2001
LOCUS Homo sapiens chromosome 19 clone LLNLR-258F7, complete sequence.
DEFINITION AC011541
ACCESSION AC011541.7 GI:14971181
VERSION HTG; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 39944)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39944)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 39944)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 20, 2001 this sequence version replaced gi:14042992.
Draft sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
Location/Qualifiers
1..39944
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLR-258F7"
ORIGIN
Query Match 68.7%; Score 20.6; DB 9; Length 39944;
Best Local Similarity 85.2%; Pred. No. 3.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 37747 TGGCGCTGAACCCAGGAGTAGCAAGAGT 37721
|||||

RESULT 27
AC023172/c 42547 bp DNA linear PRI 09-FEB-2000
LOCUS Homo sapiens chromosome 19, cosmid F22376, complete sequence.
DEFINITION AC023172
ACCESSION AC023172.1 GI:6957691
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

repeat_region 18249..18321
/rpt family=" (CAGAGA)n"
repeat_region 18322..18594
/rpt family="AluSc"
repeat_region 18595..18619
/rpt family=" (TAA)n"
repeat_region 18624..18720
/rpt family=" (CAGAGA)n"
repeat_region complement(18961..19259)
/rpt family="AluJo"
repeat_region 19395..19577
/rpt family="MIR"
repeat_region 20033..20221
/rpt family="GA-rich"
repeat_region complement(20268..20434)
/rpt family="AluSc"
repeat_region complement(20435..20733)
/rpt family="AluSc"
repeat_region complement(21020..21241)
/rpt family="AluSg/x"
repeat_region complement(21270..21432)
/rpt family="MERSA"
repeat_region 21675..21855
/rpt family="GA-rich"
repeat_region 22115..22440
/rpt family="G-rich"
repeat_region 22438..22646
/rpt family=" (CAGAGA)n"
repeat_region complement(23201..23211)
/rpt family="L1PB3"
repeat_region complement(23212..23463)
/rpt family="AluY"
repeat_region complement(23464..23720)
/rpt family="L1PB3"
repeat_region 23736..23851
/rpt family=" (TTCA)n"
repeat_region 23911..23958
/rpt family=" (TTCA)n"
repeat_region 24033..24325
/rpt family=" (TTCA)n"
repeat_region 24326..24351
/rpt family=" (ATTG)n"
repeat_region 24352..24384
/rpt family=" (TTCA)n"
repeat_region 24543..24568
/rpt family=" (TG)n"
repeat_region complement(24763..25059)
Query Match 68.7%; Score 20.6; DB 9; Length 42547;
Best Local Similarity 85.2%; Pred. No. 3.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 TGGCGCAGACGAGGTAGCACCACAT 27
Db 4183 TGGCGCTGACGAGGTAGCAGAGT 4157
RESULT 28
AC092374/c AC092374 150962 bp DNA linear PRI 28-FEB-2003
LOCUS Homo sapiens chromosome 16 clone RP11-63716, complete sequence.
DEFINITION AC092374 AC022789
ACCESSION AC092374.4 GI:28603936
VERSION AC092374.4
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150962)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 150962)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 150962)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 150962)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 28, 2003 this sequence version replaced gi:19745005.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
Location/Qualifiers
source 1..150962
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="16"
/clone="RP11-63718"
ORIGIN
Query Match 68.7%; Score 20.6; DB 9; Length 150962;
Best Local Similarity 85.2%; Pred. No. 3.8e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 GGGCGAGAACGAGGTAGCACCACATG 28
Db 128079 GGGCGAARACGAGGTAAACCAATG 128053
RESULT 29
AC145155/c AC145155 164831 bp DNA linear HTG 10-JUN-2003
LOCUS Homo sapiens chromosome 5 clone RP11-99N22, WORKING DRAFT SEQUENCE.
DEFINITION AC145155 GI:31560162
ACCESSION AC145155
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164831)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164831)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 444715
Center clone name: RPCI-11_99N22

Summary Statistics
Consensus quality: 164579 bases at least Q40

Consensus quality: 164754 bases at least Q30
 Consensus quality: 164829 bases at least Q20
 Estimated insert size: 175000; agarose-fp estimation
 Estimated insert size: 164831; sum-of-contigs estimation
 Quality coverage: 11.07 in Q20 bases; agarose-fp estimation
 Quality coverage: 11.75 in Q20 bases; sum-of-contigs estimation
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 164831: contig of 164831 bp in length.

FEATURES

source

Location/Qualifiers
 1. .164831
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-99N22"
 /clone_lib="RP11-99N22"
 /clone_lib="RP11-99N22"

ORIGIN

Query Match 58.7%; Score 20.6; DB 2; Length 164831;
 Best Local Similarity 85.2%; Pred. No. 3.9e-02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGCAGACACAGGAGTAGCACCATTG 28

Db 12235 GCGGCAAAACAGGAGTAACCAATTG 12209

RESULT 30

AC117614/c

LOCUS

DEFINITION Mus musculus chromosome 5, clone RP23-110C17, complete sequence.

AC117614

AC117614.12 GI:38142550

HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 222809)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome 5, Clone RP23-110C17

Unpublished

2 (bases 1 to 222809)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,

Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,

Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigliio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 222809)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (11-OCT-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 222809)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., MacLean, C.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (01-NOV-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 1, 2003 this sequence version replaced gi:37651805.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

FEATURES

source

Location/Qualifiers
 1. .222809
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="5"
 /map="5"
 /clone="RP23-110C17"
 /clone_lib="RP11-99N22" Female Mouse BAC"
 1. .11631

misc_feature

```
/note="wgs_end_extension
clone_end:SP6"
repeat_region 1648..1729
repeat_region /rpt_family="MTE"
repeat_region 2408..2619
repeat_region /rpt_family="B3"
repeat_region 3221..3241
repeat_region /rpt_family="AT_rich"
repeat_region 3787..3816
repeat_region /rpt_family="(TGG)n"
repeat_region 4265..4344
repeat_region /rpt_family="(TCCA)n"
repeat_region complement(5866..5984)
repeat_region /rpt_family="ID_B1"
repeat_region 6022..6093
repeat_region /rpt_family="B1_MM"
repeat_region 6181..6309
repeat_region /rpt_family="B1_MM"
repeat_region 6310..6363
repeat_region /rpt_family="(CAAAAA)n"
repeat_region 6557..6950
repeat_region /rpt_family="MTD"
repeat_region 6963..7144
repeat_region /rpt_family="LIMB7"
repeat_region 7150..7357
repeat_region /rpt_family="B3"
repeat_region complement(7378..7524)
repeat_region /rpt_family="B1_MM"
repeat_region 7552..7677
repeat_region /rpt_family="RSINE1"
repeat_region 7692..7721
repeat_region /rpt_family="(CAAT)n"
repeat_region 7733..7948
repeat_region /rpt_family="LIMB7"
repeat_region complement(8580..9057)
repeat_region /rpt_family="IAPLTR2_MM"
repeat_region 10093..10284
repeat_region /rpt_family="B3"
repeat_region 10291..10402
repeat_region /rpt_family="PB1D9"
repeat_region 10413..10451
repeat_region /note="30 qual SINGL region"
repeat_region 11112..11213
repeat_region /rpt_family="PB1D7"
repeat_region 11215..11259
repeat_region /rpt_family="GA-rich"
repeat_region 11265..11382
repeat_region /rpt_family="B1_MM"
repeat_region 11383..11413
repeat_region /rpt_family="(CAAA)n"
repeat_region 11427..11976
repeat_region /rpt_family="LIM4"
misc_feature 11632..11637
misc_feature /note="clone boundary
clone_end:SP6
site:ECORI"
repeat_region 13247..13399
repeat_region /rpt_family="MER2"
repeat_region 13662..13761
repeat_region /rpt_family="PB1D7"
repeat_region 14295..14411
repeat_region /rpt_family="B1_MM"
repeat_region 14992..15035
repeat_region /rpt_family="B2_MM2"
repeat_region 15023..15154
repeat_region /rpt_family="B3"
repeat_region 15528..15557
repeat_region /rpt_family="(GGA)n"
repeat_region 16277..16299
repeat_region /rpt_family="AT_rich"
repeat_region complement(16384..16505)
repeat_region /rpt_family="B3A"
repeat_region 16593..16770
```

```
repeat_region /rpt_family="B2_MM2"
repeat_region 16799..16918
repeat_region /rpt_family="B2_MM2"
repeat_region complement(17278..17316)
repeat_region /rpt_family="MER5B"
repeat_region 17385..17487
repeat_region /rpt_family="B1_MM"
repeat_region 17686..17792
repeat_region /rpt_family="B1_MM"
repeat_region complement(17783..17847)
repeat_region /rpt_family="PB1D10"
repeat_region 18153..18172
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(18173..18351)
repeat_region /rpt_family="B2_MM1"
repeat_region complement(19078..19471)
repeat_region /rpt_family="MTA"

Query Match 68.78; Score 20.6; DB 10; Length 222809;
Best Local Similarity 85.24; Pred.No.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCGAGGAGTACCAAT 27
Db 118038 TGGCTCAGAACCATGAGTAACCCCAAT 118012
|||||
|||||
```

Search completed: June 20, 2004, 11:44:43
Job time : 836.603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 172.381 Seconds
(without alignment)
739.327 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgacagaccaggtagcaccacatgag 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | Match | Length | DB ID | Description |
|------------|-------|---------|--------|--------|----------|--------------------|
| 1 | 30 | 100.0 | 1584 | 4 | AH51956 | Aah51956 Mycobacte |
| 2 | 30 | 100.0 | 110000 | 4 | Aai99682 | Aai99682 Mycobacte |
| 3 | 30 | 100.0 | 110000 | 4 | Aai99683 | Aai99683 Mycobacte |
| 4 | 20.6 | 68.7 | 1476 | 2 | AAG87723 | Aag87723 Human aux |
| 5 | 20.6 | 68.7 | 1476 | 2 | AAT28389 | Aat28389 Human cyt |
| 6 | 20.6 | 68.7 | 1476 | 2 | AAT17417 | Aat17417 Human der |
| 7 | 20.6 | 68.7 | 2915 | 2 | AAH11662 | AAh11662 Human CYP |
| 8 | 20.6 | 68.7 | 3045 | 2 | AAZ19787 | Aaz19787 Human cyt |
| 9 | 20.6 | 68.7 | 3045 | 2 | AAZ19787 | Aaz19787 Human cyt |
| 10 | 20.6 | 68.7 | 3045 | 6 | ABL64908 | Abi64908 Lung canc |
| 11 | 20.6 | 68.7 | 3045 | 6 | ABT10310 | Abt10310 Human bre |
| 12 | 20.6 | 68.7 | 3045 | 6 | ABN95821 | Abn95821 Gene #231 |
| 13 | 20.6 | 68.7 | 4887 | 9 | ADD29818 | Add29818 Human tum |
| 14 | 20.4 | 68.0 | 41 | 4 | AAH11651 | AAh11651 Human CYP |
| 15 | 19.6 | 65.3 | 590 | 4 | AAH69361 | Aah69361 Human cer |
| 16 | 19.6 | 65.3 | 649 | 4 | AAH72227 | Aah72227 Human cer |
| 17 | 19.6 | 65.3 | 778 | 4 | AAH71164 | Aah71164 Human cer |
| 18 | 19.6 | 65.3 | 2420 | 4 | AAH72859 | Aah72859 Human cer |
| 19 | 19.6 | 65.3 | 2420 | 5 | ABV25077 | Abv25077 Human pro |
| 20 | 19.6 | 65.3 | 2427 | 4 | ABL27011 | Abi27011 Drosophil |
| 21 | 19.6 | 65.3 | 6200 | 4 | ABL27010 | Abi27010 Drosophil |
| 22 | 19.4 | 64.7 | 537 | 4 | ABA61254 | AbA61254 Human foe |
| 23 | 19.4 | 64.7 | 537 | 4 | AAI41155 | AAi41155 Probe #98 |

| | | | | | | |
|----|------|------|--------|---|-------------|--------------------|
| 24 | 19.4 | 64.7 | 537 | 4 | ABA29093 | AbA29093 Probe #75 |
| 25 | 19.4 | 64.7 | 537 | 4 | AAK35442 | Aak35442 Human bon |
| 26 | 19.4 | 64.7 | 537 | 4 | AAK09550 | Aak09550 Human bra |
| 27 | 19.4 | 64.7 | 537 | 4 | ABS35177 | AbS35177 Human liv |
| 28 | 19.4 | 64.7 | 537 | 6 | ABS09801 | AbS09801 Human gen |
| 29 | 19.2 | 64.0 | 21706 | 2 | AAV52169 | Aav52169 Streptoco |
| 30 | 19.2 | 64.0 | 110000 | 7 | ABS6454_14 | Continuation (15 o |
| 31 | 19 | 63.3 | 280 | 3 | AAC16260 | Aac16260 Human sec |
| 32 | 19 | 63.3 | 393 | 8 | ACH17812 | Ach17812 Human adu |
| 33 | 19 | 63.3 | 637 | 6 | ABS72349 | AbS72349 Human gen |
| 34 | 19 | 63.3 | 843 | 6 | ABX97067 | AbX97067 Human NOV |
| 35 | 19 | 63.3 | 1174 | 7 | ABX05033 | AbX05033 Human nov |
| 36 | 19 | 63.3 | 40178 | 9 | ADE43862 | Ade43862 Human LIP |
| 37 | 19 | 63.3 | 40178 | 9 | ADE43863 | Ade43863 Polymorph |
| 38 | 19 | 63.3 | 95269 | 6 | ABQ67195_4 | Continuation (14 o |
| 39 | 19 | 63.3 | 110000 | 6 | ABQ69245_13 | Continuation (15 o |
| 40 | 18.8 | 62.7 | 216 | 4 | ABL27759 | AbL27759 Drosophil |
| 41 | 18.8 | 62.7 | 289 | 4 | AAI26095 | Aai26095 Human bre |
| 42 | 18.8 | 62.7 | 328 | 4 | AAI26095 | Aai26095 Human bre |
| 43 | 18.8 | 62.7 | 328 | 4 | AAI18165 | Aai18165 Human bre |
| 44 | 18.8 | 62.7 | 365 | 4 | AAI22557 | Aai22557 Human bre |
| 45 | 18.8 | 62.7 | 366 | 4 | AAI13845 | Aai13845 Human bre |
| 46 | 18.8 | 62.7 | 366 | 4 | AAI22588 | Aai22588 Human bre |
| 47 | 18.8 | 62.7 | 369 | 4 | AAI22711 | Aai22711 Human bre |
| 48 | 18.8 | 62.7 | 372 | 4 | AAI13986 | Aai13986 Human bre |
| 49 | 18.8 | 62.7 | 377 | 4 | AAI13650 | Aai13650 Human bre |
| 50 | 18.8 | 62.7 | 381 | 4 | AAI13650 | Aai13650 Human bre |
| 51 | 18.8 | 62.7 | 438 | 4 | AAI13688 | Aai13688 Human bre |
| 52 | 18.8 | 62.7 | 439 | 4 | AAI13758 | Aai13758 Human bre |
| 53 | 18.8 | 62.7 | 575 | 4 | AAI22625 | Aai22625 Human bre |
| 54 | 18.8 | 62.7 | 1015 | 3 | AAC59665 | Aac59665 Human sec |
| 55 | 18.8 | 62.7 | 1069 | 7 | AAD51043 | Aad51043 Human DME |
| 56 | 18.8 | 62.7 | 1143 | 2 | AAV42730 | Aav42730 Human cyt |
| 57 | 18.8 | 62.7 | 1406 | 3 | AAC59652 | Aac59652 Human sec |
| 58 | 18.8 | 62.7 | 1440 | 7 | ABX53820 | AbX53820 Human CDN |
| 59 | 18.8 | 62.7 | 1500 | 6 | ABQ76539 | AbQ76539 C. albica |
| 60 | 18.8 | 62.7 | 1824 | 2 | AAT89089 | Aat89089 Avian 3-h |
| 61 | 18.8 | 62.7 | 2216 | 4 | ABL27758 | AbL27758 Drosophil |
| 62 | 18.8 | 62.7 | 2234 | 4 | ABL24922 | AbL24922 Drosophil |
| 63 | 18.8 | 62.7 | 2815 | 4 | ABL09420 | AbL09420 Drosophil |
| 64 | 18.8 | 62.7 | 2907 | 6 | ABL55916 | AbL55916 Lung canc |
| 65 | 18.8 | 62.7 | 2907 | 6 | ABL55008 | AbL55008 Lung canc |
| 66 | 18.8 | 62.7 | 2907 | 6 | ABL5513 | AbL5513 Lung canc |
| 67 | 18.8 | 62.7 | 2907 | 6 | ABN95820 | AbN95820 Gene #231 |
| 68 | 18.8 | 62.7 | 3704 | 4 | ABL09414 | AbL09414 Drosophil |
| 69 | 18.8 | 62.7 | 4321 | 4 | ABL18634 | AbL18634 Drosophil |
| 70 | 18.6 | 62.0 | 290 | 7 | ABX88273 | AbX88273 Corn ear- |
| 71 | 18.6 | 62.0 | 936 | 9 | ADC33952 | Adc33952 DNA seque |
| 72 | 18.6 | 62.0 | 1283 | 2 | AAK60284 | Aak60284 Nucleic a |
| 73 | 18.6 | 62.0 | 1406 | 4 | ABL06013 | AbL06013 Drosophil |
| 74 | 18.6 | 62.0 | 3184 | 4 | ABL18159 | AbL18159 Drosophil |
| 75 | 18.6 | 62.0 | 3480 | 4 | ABL06012 | AbL06012 Drosophil |
| 76 | 18.6 | 62.0 | 7138 | 4 | ABL18158 | AbL18158 Drosophil |
| 77 | 18.4 | 61.3 | 228 | 4 | ABL25165 | AbL25165 Drosophil |
| 78 | 18.4 | 61.3 | 333 | 3 | AAC57133 | Aac57133 Pinus rad |
| 79 | 18.4 | 61.3 | 394 | 3 | AAC07594 | Aac07594 Human sec |
| 80 | 18.4 | 61.3 | 546 | 6 | ABX55262 | AbX55262 Human col |
| 81 | 18.4 | 61.3 | 627 | 4 | AAI10291 | AAi10291 Drosophil |
| 82 | 18.4 | 61.3 | 729 | 4 | AAK89066 | Aak89066 Human dig |
| 83 | 18.4 | 61.3 | 729 | 4 | AAI57623 | AAi57623 Human col |
| 84 | 18.4 | 61.3 | 729 | 6 | ABS98800 | AbS98800 Genomic D |
| 85 | 18.4 | 61.3 | 729 | 9 | ADB92953 | AdB92953 Human ova |
| 86 | 18.4 | 61.3 | 752 | 6 | ABQ56190 | AbQ56190 Human col |
| 87 | 18.4 | 61.3 | 778 | 4 | AAH33523 | Aah33523 Human col |
| 88 | 18.4 | 61.3 | 864 | 6 | ABS99499 | AbS99499 Enterococ |
| 89 | 18.4 | 61.3 | 864 | 6 | ABS99499 | AbS99499 Enterococ |
| 90 | 18.4 | 61.3 | 888 | 4 | AAK89065 | Aak89065 Human dig |
| 91 | 18.4 | 61.3 | 888 | 4 | AAI57622 | AAi57622 Human col |
| 92 | 18.4 | 61.3 | 888 | 6 | ABS99799 | AbS99799 Genomic D |
| 93 | 18.4 | 61.3 | 888 | 9 | ADB29552 | AdB29552 Human col |
| 94 | 18.4 | 61.3 | 910 | 4 | AAI92558 | AAi92558 Human pol |
| 95 | 18.4 | 61.3 | 931 | 5 | AAH23841 | Aah23841 Human tra |
| 96 | 18.4 | 61.3 | 1059 | 6 | ABZ14011 | AbZ14011 Arabidops |

| | | | | | | | | | | | | | | |
|-----|------|------|--------|---|-----------|--------------------|-------|----|------|------|---|----------|----------|-----------|
| 97 | 18.4 | 61.3 | 1059 | 7 | ADA68115 | Arabidops | 170 | 18 | 60.0 | 1203 | 7 | ADA69469 | Ada69469 | Rice gene |
| 98 | 18.4 | 61.3 | 1246 | 4 | AD05827 | Arabidops | 171 | 18 | 60.0 | 1267 | 6 | ABN98257 | ABN98257 | Arabidops |
| 99 | 18.4 | 61.3 | 1246 | 5 | AD06666 | A. thalia | c 172 | 18 | 60.0 | 1337 | 3 | AAC48518 | AAC48518 | Arabidops |
| 100 | 18.4 | 61.3 | 1246 | 6 | ABK65295 | Arabidops | c 173 | 18 | 60.0 | 1355 | 3 | AAC48403 | AAC48403 | Arabidops |
| 101 | 18.4 | 61.3 | 1246 | 8 | ADAI5568 | DNA encod | c 174 | 18 | 60.0 | 1377 | 3 | AAC41116 | AAC41116 | Arabidops |
| 102 | 18.4 | 61.3 | 1246 | 9 | ADAI5568 | DNA encod | c 175 | 18 | 60.0 | 1447 | 3 | AAC41073 | AAC41073 | Arabidops |
| 103 | 18.4 | 61.3 | 1246 | 9 | ADAI5568 | Thalecres | c 176 | 18 | 60.0 | 1452 | 3 | AAC41073 | AAC41073 | Arabidops |
| 104 | 18.4 | 61.3 | 1246 | 9 | ADAI5568 | Plant yle | c 177 | 18 | 60.0 | 1512 | 8 | AD09996 | AD09996 | Alloccoc |
| 105 | 18.4 | 61.3 | 1246 | 9 | ADAI5568 | Plant yle | c 178 | 18 | 60.0 | 1512 | 8 | AD09996 | AD09996 | Alloccoc |
| 106 | 18.4 | 61.3 | 1311 | 2 | AAC34833 | Arabidops | 179 | 18 | 60.0 | 1932 | 7 | ACD01710 | ACD01710 | DNA sequ |
| 107 | 18.4 | 61.3 | 1373 | 2 | AAC34833 | Arabidops | 180 | 18 | 60.0 | 1932 | 7 | ACD01710 | ACD01710 | DNA sequ |
| 108 | 18.4 | 61.3 | 1884 | 3 | AAC4542 | Arabidops | 181 | 18 | 60.0 | 1932 | 7 | ACD01710 | ACD01710 | DNA sequ |
| 109 | 18.4 | 61.3 | 2100 | 2 | AAV11356 | Triochoder | 182 | 18 | 60.0 | 1932 | 7 | ACD01682 | ACD01682 | DNA sequ |
| 110 | 18.4 | 61.3 | 2288 | 4 | AB125164 | Drosophill | 183 | 18 | 60.0 | 1932 | 7 | ACD01682 | ACD01682 | DNA sequ |
| 111 | 18.4 | 61.3 | 2471 | 2 | AAK35516 | CDNA encod | 184 | 18 | 60.0 | 1932 | 7 | ACD01689 | ACD01689 | DNA sequ |
| 112 | 18.4 | 61.3 | 2855 | 6 | ABK35764 | CDNA sequ | 185 | 18 | 60.0 | 1932 | 7 | ACD01713 | ACD01713 | DNA sequ |
| 113 | 18.4 | 61.3 | 3443 | 4 | AB110290 | Drosophill | 186 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 114 | 18.4 | 61.3 | 3859 | 7 | ABT16586 | Ethylene | 187 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 115 | 18.4 | 61.3 | 4581 | 5 | AA669032 | DNA encod | 188 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 116 | 18.4 | 61.3 | 7058 | 7 | AB224705 | Human cel | 189 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 117 | 18.4 | 61.3 | 7931 | 4 | AAK85096 | Human imm | 190 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 118 | 18.4 | 61.3 | 9731 | 4 | AAK72933 | Human imm | 191 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 119 | 18.4 | 61.3 | 9733 | 4 | AAK85097 | Human imm | 192 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 120 | 18.4 | 61.3 | 9733 | 4 | AAK85097 | Human imm | 193 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 121 | 18.4 | 61.3 | 9733 | 4 | AAK85097 | Human imm | 194 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 122 | 18.2 | 60.7 | 518 | 3 | AAK04044 | Human sec | 195 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 123 | 18.2 | 60.7 | 894 | 3 | AAK359565 | Human sec | 196 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 124 | 18.2 | 60.7 | 1239 | 9 | ACF69207 | Photocarb | c 197 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 125 | 18.2 | 60.7 | 2306 | 9 | ADDA4565 | Human gen | 198 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 126 | 18.2 | 60.7 | 5424 | 5 | AAK73915 | DNA encod | 199 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 127 | 18.2 | 60.7 | 5956 | 6 | AAK73915 | DNA encod | 200 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 128 | 18.2 | 60.7 | 7643 | 6 | ABQ99421 | Human cod | 201 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 129 | 18.2 | 60.7 | 7931 | 9 | ABE71077 | Human cod | 202 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 130 | 18.2 | 60.7 | 10620 | 4 | AB117332 | Drosophill | 203 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 131 | 18.2 | 60.7 | 12748 | 4 | AB117332 | Drosophill | 204 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 132 | 18.2 | 60.7 | 110000 | 4 | AB119662 | Continuation (37 o | 205 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 133 | 18.2 | 60.7 | 110000 | 4 | AB119662 | Continuation (37 o | 206 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 134 | 18.2 | 60.7 | 110000 | 7 | ACF67367 | Continuation (22 o | 207 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 135 | 18.2 | 60.7 | 243072 | 7 | ACF67367 | Continuation (22 o | 208 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 136 | 18.2 | 60.7 | 406 | 6 | ABK27687 | Photocarb | 209 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 137 | 18.2 | 60.7 | 413 | 7 | ABK42725 | Human col | 210 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 138 | 18.2 | 60.7 | 413 | 7 | ABK42725 | Bovine ES | 211 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 139 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 212 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 140 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 213 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 141 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 214 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 142 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 215 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 143 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 216 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 144 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 217 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 145 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 218 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 146 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 219 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 147 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 220 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 148 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 221 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 149 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 222 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 150 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 223 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 151 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 224 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 152 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 225 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 153 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 226 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 154 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 227 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 155 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 228 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 156 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 229 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 157 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 230 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 158 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 231 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 159 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 232 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 160 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 233 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 161 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 234 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 162 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 235 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 163 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 236 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 164 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 237 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 165 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 238 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 166 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 239 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 167 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 240 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 168 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 241 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |
| 169 | 18.2 | 60.7 | 427 | 4 | AA110161 | Bovine ES | 242 | 18 | 60.0 | 1932 | 9 | ADB71121 | ADB71121 | DNA sequ |

Drug target; growth; organism viability; characterisation; ds.
Mycobacterium tuberculosis.
17-MAY-2001.
13-NOV-2000; 2000WO-US031152.
12-NOV-1999; 99US-0165086P.
12-NOV-1999; 99US-0165124P.
01-FEB-2000; 2000US-0179531P.
(REGC) UNIV CALIFORNIA.
Eisenberg D, Rotstein SH, Marcotte EM;
WPI; 2001-329193/34.
P-PSDB; AAG81105.
Identifying nucleotide or polypeptide sequence for use as drug target,
involves providing algorithm that analyzes a functional relationship
between nucleotide or polypeptide sequences, and comparing the sequences.
Disclosure; Page 53; 207pp; English.
This invention relates to a method for identifying a nucleotide or
polypeptide sequence that may be a drug target, or essential for growth
or viability of an organism. Polynucleotide sequences AAG81096 - AAG81241, Mycobacterium
represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
tuberculosis proteins which are potential drug targets. The DNA and
protein sequences are used to illustrate the method of the invention. The
method involves providing an unknown nucleotide or polypeptide sequence,
and comparing it to a number of sequences along with at least one
algorithm capable of analysing a functional relationship between
nucleotide and polypeptide sequences. The method is useful for
characterising the function of nucleic acids and polypeptides that may be
useful as a target for a drug or essential for the growth or viability of
an organism
Sequence 1584 BP; 299 A; 539 C; 512 G; 234 T; 0 U; 0 Other;
Query Match 100.0%; Score 30; DB 4; Length 1584;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Ct 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30
Db 441 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 470
RESULT 2
AAI99682_00/c
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000

AAI99682_00/c
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000

ALIGNMENTS

RESULT 1
AAH51956
ID AAH51956 standard; DNA; 1584 BP.
AC AAH51956;
DT 04-SEP-2001 (first entry)
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 10.
XX

WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529
ID AAI99682 standard; DNA; 4411529 BP.
XX
AC AAI99682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-00103840.
XX
PR 24-JUN-1998; 98US-00103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
PX WPI; 2001-647261/74.
XX
DR Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC 1551
PT and H37Rv differ.
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen.
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly

CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1
XX
SQ Sequence 4411529 BP; 758565A; 1449983C; 1444602G; 758379T; 0U; 0Other;
Query Match 100.0%; Score 30; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGCGGAGAACCGAGTAGCACCACATGAG 30
DB 25004 TGGCGGAGAACCGAGTAGCACCACATGAG 24975

RESULT 3
AAI99683_00/c
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765
ID AAI99683 standard; DNA; 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.

XX PN US6294328-B1.
XX PD 25-SEP-2001.
XX PF 24-JUN-1998; 98US-00103840.
XX PR 24-JUN-1998; 98US-00103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX DR WPI; 2001-647261/74.
XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC 1551
PT and H37Rv differ.
XX PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docID=6294328B1
XX SQ Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890Other;
Query Match 100.0%; Score 30; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACGAGGTAGCACCAATGAG 30
DB 24986 TGGCGCAGAACGAGGTAGCACCAATGAG 24957
RESULT 4
AAQ87723/c
ID AAQ87723 standard; cDNA; 1476 BP.
XX AC AAQ87723;
XX DT 25-MAR-2003 (revised)
XX DT 15-NOV-1995 (first entry)
XX DE Human auxillary cytochrome P450 species 2B6 coding region.
XX KW Human cytochrome P450; amplification; PCR; primer; expression vector;
KW Yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW carcinogen; mutagen; liver metabolism; ds.
XX OS Homo sapiens.
XX PN EP644267-A2.
XX PD 22-MAR-1995.
XX PF 20-JUL-1994; 94EP-00111298.
XX PR 20-JUL-1993; 93JP-00201120.
XX PR 21-JUL-1993; 93JP-00180246.
XX PR 30-JUL-1993; 93JP-00208279.

PA (SUMO) SUMITOMO CHEM CO LTD.
XX Hayashi K, Sakaki T, Yabuseki Y, Komai K, Kaneko H, Nakatsuka I;
XX WPI; 1995-116991/16.
XX DR P-PSDB; AAR72369.
XX PT Evaluation of safety of a chemical cpd. - using recombinant yeast
PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase.
XX PS Example; Page 57-59; 124pp; English.
XX CC The nucleotide sequence of the cDNA coding region for the human auxillary
CC cytochrome P450 species 2B6. The gene encodes a protein of 491 amino
CC acids. The cDNA was amplified by PCR using the primers AAQ87755-8. The
CC product was cloned into the yeast expression vectors pAAH8N or pAHR8 to
CC produce the vectors p2B6 for the expression of the cytochrome P450 alone
CC or p2B6R for co-expression with the yeast NADPH-P450 reductase. The
CC vectors are used in a method for evaluating the safety of a chemical
CC compound by reacting the chemical compound with recombinantly produced
CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
CC 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species and variants
CC or in cell extracts, and analysing the resulting metabolite to assess the
CC safety of the chemical compound. The method is useful for determining
CC whether the chemical compound, or its metabolite, will be converted into
CC a carcinogenic or mutagenic form through metabolism in the liver.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 1476 BP; 351 A; 453 C; 338 G; 334 T; 0 U; 0 Other;
Query Match 68.7%; Score 20.6; DB 2; Length 1476;
Best Local Similarity 85.2%; Pred. No. 1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACGAGGTAGCACCAAT 27
DB 68 TGGCGCTGAACGAGGAGTAGCAAGT 42
RESULT 5
AAT28389/c
ID AAT28389 standard; DNA; 1476 BP.
XX AC AAT28389;
XX DT 11-OCT-1996 (first entry)
XX DE Human cytochrome P450 molecular species 2B6 gene.
XX KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
KW evaluation; safety; fusion protein; metabolite; detoxification;
KW carcinogenic; ds.
XX OS Homo sapiens.
XX PN JP08056695-A.
XX PD 05-MAR-1996.
XX PF 15-JUL-1994; 94JP-00164184.
XX PR 20-JUL-1993; 93JP-00201120.
XX PR 30-JUL-1993; 93JP-00208279.
XX PR 17-JUN-1994; 94JP-00136053.
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX DR WPI; 1996-182311/19.
XX DR P-PSDB; AAR93176.
XX PT Novel method for the evaluation of the safety of a cpd. - using a human

PT cytochrome P450 and yeast NADPH reductase to determine whether the
 PT analyte cpd. is detoxified or metabolised to a carcinogen.
 XX
 PS Example 1; Page 37-39; 74pp; Japanese.
 XX This is the nucleotide sequence of the human cytochrome P450 molecular
 CC species 2B6 gene which encodes a protein of 491 amino acids. The 1.5 kb
 CC fragment was amplified from a human liver derived cDNA library using
 CC primers AAT2645-6. The prod. was cloned into the yeast expression vector
 CC pAAH5N to generate plasmid p2B6 for prodn. of the cytochrome only or into
 CC the vector pAHR to generate the plasmid p2B6R for co-prodn. with the
 CC yeast NADPH-P450 reductase. The sequence is placed under control of the
 CC yeast ADH gene promoter and terminator. The vectors are used in a method
 CC for evaluating the safety of a cpd. by reacting the test cpd. with
 CC recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),
 CC 2C9 (28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants (AAT28384-
 CC 98) together with yeast NADPH-P450 reductase (either as a fused protein
 CC or as a cell extract) and analysing the resultant metabolite. The cpd. is
 CC considered "safe" if it is detoxified or not rendered carcinogenic or
 CC "unsafe" if it is not detoxified or is metabolised to a carcinogenic cpd
 XX
 SQ Sequence 1476 BP; 351 A; 453 C; 338 G; 334 T; 0 U; 0 Other;
 Query Match 68.7%; Score 20.6; DB 2; Length 1476;
 Best Local Similarity 85.2%; Pred. No. 1e+02; Mismatches 0; Gaps 0;
 Matches 23; Conservative 0; Indels 4; Indels 0; Gaps 0;
 QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
 DB 68 TGGCGCTGAACCCAGGAGTAGCAGAGT 42
 RESULT 6
 AAT17417/C
 ID AAT17417 standard; cDNA; 1476 BP.
 AC AAT17417;
 XX
 XX 01-AUG-1996 (first entry)
 DT
 DE Human derived cytochrome P4502B6 cDNA.
 DE Human derived cytochrome; P4502B6; commercial cDNA library; yeast;
 KW transfection; recombinant production; expression vector; mammal;
 KW immunisation; sensitisation; antibody; determination; detection;
 KW non-cross reactive; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1476
 FT /*tag= a
 FT
 XX JP08027193-A.
 PN
 XX 30-JAN-1996.
 PD
 XX 13-JUL-1994; 94JP-00161548.
 PF
 XX 13-JUL-1994; 94JP-00161548.
 PR
 XX (SUMO) SUMITOMO CHEM CO LTD.
 PA
 XX WPI: 1996-136334/14.
 DR P-PSDB; AAR81468.
 XX
 XX Antibody recognising human derived cytochrome P4502B6 - allows specific
 PT detection of cytochrome P450 species in humans.
 PS
 XX Example 1; Page 10-12; 13pp; Japanese.
 XX The present sequence encodes the human derived cytochrome (HDC) P4502B6,
 CC which was obtd. from a commercial cDNA library. Yeast were transfected
 CC

CC with an expression vector contg. the HDC cDNA, cultured and then
 CC disrupted to give a microsomal fraction. The HDC was purified from the
 CC fraction, and used to immunise and sensitise a mammal. Blood was drawn
 CC from the mammal, and an anti-HDC antibody isolated. The antibody obtd.
 CC recognises HDC P4502B6, partic. at a serum dilution rate of 1:10000, and
 CC is substantially without cross reaction to other HDC P450 spp
 XX
 SQ Sequence 1476 BP; 351 A; 453 C; 338 G; 334 T; 0 U; 0 Other;
 Query Match 68.7%; Score 20.6; DB 2; Length 1476;
 Best Local Similarity 85.2%; Pred. No. 1e+02; Mismatches 0; Gaps 0;
 Matches 23; Conservative 0; Indels 4; Indels 0; Gaps 0;
 QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
 DB 68 TGGCGCTGAACCCAGGAGTAGCAGAGT 42
 RESULT 7
 AAS11662/c
 ID AAS11662 standard; DNA; 2915 BP.
 XX AC AAS11662;
 XX
 XX 24-OCT-2001 (first entry)
 DT
 XX Human CYP2B6 DNA 5' flanking sequence.
 DE
 XX CYP2B6; cytostatic; gene therapy; genotyping; cancer; metabolism; ds;
 KW human; cancer susceptibility; environmental carcinogen;
 KW 5' flanking sequence.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT primer_bind complement(1..18)
 FT /*tag= a
 FT /bound_moiety= "PCR primer CYP2B6-P2F"
 FT primer_bind complement(565..584)
 FT /*tag= b
 FT /bound_moiety= "Sequencing primer seqCYP2B6-P1F"
 FT variation replace(688,G)
 FT /*tag= c
 FT variation replace(1010,C)
 FT /*tag= d
 FT primer_bind complement(1212..1232)
 FT /*tag= e
 FT /bound_moiety= "Sequencing primer seqCYP2B6-P2F"
 FT variation replace(1280,G)
 FT /*tag= f
 FT variation replace(1715,C)
 FT /*tag= g
 FT primer_bind complement(1865..1883)
 FT /*tag= h
 FT /bound_moiety= "Sequencing primer seqCYP2B6-P3F"
 FT primer_bind 1944..1964
 FT /*tag= i
 FT /bound_moiety= "Sequencing primer seqCYP2B6-P4R"
 FT primer_bind complement(2193..2210)
 FT /*tag= j
 FT /bound_moiety= "PCR primer CYP2B6-P1F"
 FT primer_bind 2193..2210
 FT /*tag= k
 FT /bound_moiety= "PCR primer CYP2B6-P1R"
 FT variation replace(2383,C)
 FT /*tag= l
 FT exon 2455..2550
 FT /*tag= m
 FT /number= 1
 FT 2507..2526
 FT primer_bind /*tag= n
 FT /bound_moiety= "Sequencing primer seqCYP2B6-P5R"
 FT primer_bind 2898..2915

```

FT      /*tag= o
FT      /bound_moisty= "PCR primer CYP2B6-Pr"
XX      WO200159152-A2.
XX      PD 16-AUG-2001.
XX      XX
XX      PF 09-FEB-2001; 2001WO-EP001456.
XX      XX
XX      PR 09-FEB-2000; 2000EP-00102701.
XX      XX
XX      PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX      XX
XX      PI Zanger UM, Lang T;
XX      XX
XX      DR WPI; 2001-502719/55.
XX      XX
XX      PT New polynucleotide(s) of the polymorphic human CYP2B6 gene for the
XX      PT detection and treatment of disorders i.e. cancer.
XX      PS Example 2; Fig 2; 83pp; English.
XX      CC The sequence represents the 5' flanking sequence of human CYP2B6 DNA.
XX      CC CYP2B6 polynucleotides and oligonucleotides are used for specific
XX      CC detection and genotyping of CYP2B6 alleles in humans, determination of
XX      CC which is useful for the optimisation of therapies utilising CYP2B6
XX      CC substrates. Oligonucleotide sequences are useful in detection of the
XX      CC individual predisposition to several common cancers caused by
XX      CC environmental carcinogens, and diseases treated with drugs that are
XX      CC targets of the CYP2B6 gene product, whose metabolism is therefore
XX      CC dependent on CYP2B6. Cancer or susceptibility to cancer can be diagnosed
XX      CC by detecting the presence of a molecular variant of CYP2B6. From variants
XX      CC of the alleles, modulators of the activity can be developed for use in
XX      CC treatment and prevention of CYP2B6-related disorders
XX      SQ Sequence 2915 BP; 812 A; 714 C; 645 G; 744 T; 0 U; 0 Other;

Query Match      68.7%; Score 20.6; DB 4; Length 2915;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGACACGAGTAGCACCAAT 27
Db 2532 TGGCGCTGACACGAGTAGCAAGAGT 2506

RESULT 8
AAZ19787/c
ID AAZ19787 standard; cDNA; 3045 BP.
XX AC AAZ19787;
XX XX
XX DT 06-DEC-1999 (first entry)
XX XX
XX DE Human cytochrome P450 2B6 cDNA.
XX KW Cytochrome; targeting; localisation; cancer; tumour; produg; reduction;
XX KW ds.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 7..1482
XX FT /*tag= a
XX FT /product= "Human cytochrome P450 2B6 CYP2B6"
XX FT intron 1266..1446
XX FT /*tag= b
XX FT FT
XX XX WO9945127-A2.
XX XX
XX PD 10-SEP-1999.
XX XX
XX PF 05-MAR-1999; 99WO-GB000674.

FT      /*tag= o
FT      /bound_moisty= "PCR primer CYP2B6-Pr"
XX      WO200159152-A2.
XX      PD 16-AUG-2001.
XX      XX
XX      PF 09-FEB-2001; 2001WO-EP001456.
XX      XX
XX      PR 09-FEB-2000; 2000EP-00102701.
XX      XX
XX      PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX      XX
XX      PI Zanger UM, Lang T;
XX      XX
XX      DR WPI; 2001-502719/55.
XX      XX
XX      PT New polynucleotide(s) of the polymorphic human CYP2B6 gene for the
XX      PT detection and treatment of disorders i.e. cancer.
XX      PS Example 2; Fig 2; 83pp; English.
XX      CC The sequence represents the 5' flanking sequence of human CYP2B6 DNA.
XX      CC CYP2B6 polynucleotides and oligonucleotides are used for specific
XX      CC detection and genotyping of CYP2B6 alleles in humans, determination of
XX      CC which is useful for the optimisation of therapies utilising CYP2B6
XX      CC substrates. Oligonucleotide sequences are useful in detection of the
XX      CC individual predisposition to several common cancers caused by
XX      CC environmental carcinogens, and diseases treated with drugs that are
XX      CC targets of the CYP2B6 gene product, whose metabolism is therefore
XX      CC dependent on CYP2B6. Cancer or susceptibility to cancer can be diagnosed
XX      CC by detecting the presence of a molecular variant of CYP2B6. From variants
XX      CC of the alleles, modulators of the activity can be developed for use in
XX      CC treatment and prevention of CYP2B6-related disorders
XX      SQ Sequence 2915 BP; 812 A; 714 C; 645 G; 744 T; 0 U; 0 Other;

Query Match      68.7%; Score 20.6; DB 4; Length 2915;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGACACGAGTAGCACCAAT 27
Db 2532 TGGCGCTGACACGAGTAGCAAGAGT 2506

RESULT 8
AAZ19787/c
ID AAZ19787 standard; cDNA; 3045 BP.
XX AC AAZ19787;
XX XX
XX DT 06-DEC-1999 (first entry)
XX XX
XX DE Human cytochrome P450 2B6 cDNA.
XX KW Cytochrome; targeting; localisation; cancer; tumour; produg; reduction;
XX KW ds.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 7..1482
XX FT /*tag= a
XX FT /product= "Human cytochrome P450 2B6 CYP2B6"
XX FT intron 1266..1446
XX FT /*tag= b
XX FT FT
XX XX WO9945127-A2.
XX XX
XX PD 10-SEP-1999.
XX XX
XX PF 05-MAR-1999; 99WO-GB000674.

XX      06-MAR-1998; 98GB-00004841.
XX      19-AUG-1998; 98GB-00018103.
XX      29-JAN-1999; 99GB-00002081.
XX      XX
XX      PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX      XX
XX      PI Stratford IJ, Patterson AV, Kingsman SM, Kan O, Griffiths L;
XX      PI Mitrophanous K;
XX      XX
XX      DR WPI; 1999-551046/46.
XX      DR P-FSDB; AAY42295.
XX      XX
XX      PT New produg activating agent targeted to selected cells or tissues,
XX      PT particularly hypoxic cells, for treating e.g. tumors.
XX      XX
XX      PS Example 12; Fig 4; 187pp; English.
XX      CC This sequence represents human cytochrome P450 2B6 (CYP2B6) cDNA. CYP2B6,
XX      CC and/or cytochrome P450 reductase (P450R) can be used to activate produgs
XX      CC to their active form via reduction. Administration of a produg is useful
XX      CC where the active drug may be metabolised before it reaches its site of
XX      CC action or where the active drug is cytotoxic, e.g., anticancer drugs.
XX      CC Cytochrome P450 isoforms such as CYP2B6, or vectors that express them,
XX      CC can be used to treat tumours, inflammation, atherosclerosis and muscular
XX      CC dystrophy, and may also be used to treat many other conditions, e.g.,
XX      CC cerebral malaria, rheumatoid arthritis, or conditions associated with
XX      CC hypoxia, ischaemia or hypoglycemia, or to deliver antibiotics, antiviral
XX      CC agents, analgesics, anaesthetics, anti-inflammatories, antineoplastic
XX      CC agents and diagnostic agents
XX      SQ Sequence 3045 BP; 706 A; 864 C; 661 G; 814 T; 0 U; 0 Other;

Query Match      68.7%; Score 20.6; DB 2; Length 3045;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGACACGAGTAGCACCAAT 27
Db 74 TGGCGCTGACACGAGTAGCAAGAGT 48

RESULT 9
AAZ07812/c
ID AAZ07812 standard; DNA; 3045 BP.
XX AC AAZ07812;
XX XX
XX DT 23-NOV-1999 (first entry)
XX XX
XX DE Human cytochrome P450 2B6 nucleotide sequence.
XX XX
XX KW Produg; localization domain; tumor-selective antibody; cytochrome P450;
XX KW produg activating domain; modified hematopoietic stem cell; MHSC; tumor;
XX KW inflammation; atherosclerosis; muscular dystrophy; cerebral malaria;
XX KW rheumatoid arthritis; hypoxia; ischemia; hypoglycemia; human; ss.
XX XX
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 7..1482
XX FT /*tag= a
XX FT /transl_except= (pos:538..540, aa: Ile)
XX FT /transl_except= (pos:925..927, aa: Leu)
XX FT misc_feature 1277..1446
XX FT /*tag= b
XX FT FT
XX FT FT /note= "the amino acid residues encoded by this region is
XX FT not indicated in the corresponding protein"
XX XX WO9945126-A2.
XX XX
XX PD 10-SEP-1999.
XX XX

```

PF 05-MAR-1999; 99WO-GB000672.
XX
PR 06-MAR-1998; 98GB-00004841.
PR 19-AUG-1998; 98GB-00018103.
XX 29-JAN-1999; 99GB-00002081.
XX
PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Stratford IJ, Patterson AV, Kingsman SM, Kan O, Griffiths L;
PI Mitrophanous K;
PI
XX WPI: 1999-540852/45.
XX P-PSDB; AA27406.
XX
PT New prodrug activating agent targeted to selected cells or tissues,
PT particularly hypoxic cells, for treating e.g. tumors or inflammation.
XX
XX Example; Fig 4; 149pp; English.
XX
CC The invention provides a new prodrug activating agent that comprises: (i)
CC a localization domain (LD; other than a tumor-selective antibody) and a
CC prodrug activating domain (PAD); (ii) at least one nucleic acid encoding
CC a cytochrome P450 and under control of at least one constitutive or
CC inducible expression control sequence or (iii) a modified hematopoietic
CC stem cell (WHSC) containing at least one nucleic acid encoding a PAD and
CC under control of elements as in (ii). The prodrug activating agent or
CC vectors that express them, are specifically used to treat tumors, or
CC inflammation, atherosclerosis and muscular dystrophy, but may also be
CC used to treat many other conditions, e.g. cerebral malaria, rheumatoid
CC arthritis, or conditions associated with hypoxia, hypoglycemia or
CC ischemia, or to deliver antibiotics, antiviral agents, analgesics,
CC anesthetics, anti-inflammatory, antineoplastic agents and diagnostic
CC agents. LD optimize activity of PAD, e.g. by delivering it to selected
CC locations or by delivering it to neighboring cells (bystander effect),
CC and allow a reduction in dose of prodrug, and thus of systemic side-
CC effects. Nucleic acids encoding the agent may be expressed selectively in
CC hypoxic cells. The present sequence represents the human cytochrome P450
CC 2B6 nucleotide sequence
XX
SQ Sequence 3045 BP; 706 A; 864 C; 661 G; 814 T; 0 U; 0 Other;

Query Match 68.7%; Score 20.6; DB 2; Length 3045;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCGCAGACAGGAGTAGCCCAAT 27
Db 74 TGCGCCTGACAGGAGTAGCCAGAGT 48

RESULT 10
ID ABL64908/c
XX ABL64908 standard; DNA; 3045 BP.
AC ABL64908;
XX
XX 15-MAY-2002 (first entry)
XX
XX Lung cancer related gene sequence SEQ ID NO:3245.
DE
DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200194629-A2.
PN
XX 13-DEC-2001.
PD
XX 30-MAY-2001; 2001WO-US010838.
PF
XX

PR 05-JUN-2000; 2000US-0209473P.
PR 03-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 22-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235837P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, and determining a change
XX in expression of a gene of a signature gene set.
PS Claim 1; SEQ ID NO 3245; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
XX neoplastic agent. The method involves exposing cells to a chemical agent
XX to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening an
XX anti-neoplastic agent, and can be used for producing a product which is
XX the data collected with respect to the anti-neoplastic agent as a result
XX of M1, and the data is sufficient to convey the chemical structure and/or
XX properties of the agent. M1 can be used in the treatment of cancer such
XX as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
XX prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
XX cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
XX cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
XX tumour

```
XX SQ Sequence 3045 BP; 706 A; 866 C; 658 G; 815 T; 0 U; 0 Other;
Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
||||| ||||||| ||||||| ||||||| |||||||
Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 11
ABT10910/c
ID ABT10910 standard; cDNA; 3045 BP.
XX AC ABT10910;
XX DT 04-DEC-2002 (first entry)
XX DE Human breast cancer associated coding sequence SEQ ID NO: 1044.
XX KW Human; breast specific gene; breast cancer; differential expression;
XX KW cytostatic; gene therapy; gene; ss.
XX OS Homo sapiens.
XX PN WO200259271-A2.
XX PD 01-AUG-2002.
XX PF 25-JAN-2002; 2002WO-US002176.
XX PR 25-JAN-2001; 2001US-0263757P.
XX PR 25-APR-2001; 2001US-0286090P.
XX PR 23-MAY-2001; 2001US-0282517P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Orr MS, Nation M, Diggins JC, Zeng W;
XX WPI; 2002-674803/72.
XX PT Diagnosing breast cancer in a patient comprises detecting the level of
XX PT gene expression in cell or tissue samples, where a differential gene
XX PT expression is indicative of breast cancer.
XX PS Claim 1; SEQ ID NO 1044; 260pp + Sequence Listing; English.
XX CC The present invention relates to methods of diagnosing breast cancer in a
XX CC patient, which comprise detecting the level of expression in a tissue
XX CC sample of two or more genes selected from those shown in ABT09867-
XX CC ABT1112, where a differential expression of the genes indicates breast
XX CC cancer. The methods are useful in diagnosing, treating, detecting the
XX CC progression, and in monitoring treatment of breast cancer in patients.
XX CC The methods are also useful as a screening tool for agents that modulate
XX CC the onset or progression of breast cancer. The breast cancer genes may be
XX CC used as diagnostic markers for the prediction or identification of the
XX CC malignant state of breast tissue, for confirming the type and progression
XX CC of cancer, and for drug screening and assays. The present sequence is a
XX CC coding sequence of the invention. Note: The sequence data for this patent
XX CC did not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 3045 BP; 706 A; 866 C; 658 G; 815 T; 0 U; 0 Other;
Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
||||| ||||||| ||||||| ||||||| |||||||
Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 12
ABN95821/c
ID ABN95821 standard; DNA; 3045 BP.
XX AC ABN95821;
XX DT 13-AUG-2002 (first entry)
XX DE Gene #2319 used to diagnose liver cancer.
XX KW Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX KW metastatic liver tumour; cytostatic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US030589.
XX PR 02-OCT-2000; 2000US-0237054P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX PT Diagnosing and detecting the progression of liver cancer, hepatocellular
XX PT carcinoma or metastatic liver tumor in a patient, involves detecting the
XX PT level of expression of two or more genes in a liver tissue sample.
XX PS Claim 1; SEQ ID NO 2319; 298pp; English.
XX CC The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatotropic, and
XX CC cytostatic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX CC for this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 3045 BP; 706 A; 866 C; 658 G; 815 T; 0 U; 0 Other;
Query Match 68.7%; Score 20.6; DB 6; Length 3045;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
||||| ||||||| ||||||| ||||||| |||||||
Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 13
ADD29818/c
ID ADD29818 standard; mRNA; 4887 BP.
XX AC ADD29818;
XX DT 15-JAN-2004 (first entry)
```

XX DE Human tumour suppressor mRNA SEQ ID NO:304.
XX DE ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.
XX KW Homo sapiens.
XX OS WO2003058201-A2.
XX PN 17-JUL-2003.
XX PD 31-DEC-2002; 2002WO-US041825.
XX PF 31-DEC-2001; 2001US-0345317P.
XX PR (QUAR-) QUARK BIOTECH INC.
XX PA (CLEV-) CLEVELAND CLINIC FOUND.
XX PI Feinstein E, Gudkov AV;
XX PI WPI; 2003-598393/56.
XX DR Diagnosing cancer comprises determining the polypeptide or polynucleotide
XX PT levels e.g., hepatic lipase, in a sample from a subject, where a higher
XX PT level compared to that in a subject free of cancer is indicative of
XX PT cancer.
XX PS Disclosure; SEQ ID NO 304; 272pp; English.
XX SQ The invention relates to a novel method for diagnosing a cancer in a
CC subject. The method comprises determining, in a sample from the subject,
CC the level of at least one polypeptide, where a higher level of the
CC polypeptide compared to the level of the polypeptide in a subject free of
CC cancer is indicative of cancer. The polypeptide is selected from any of
CC the polypeptides encoded by the polynucleotides listed in the
CC specification and polypeptides which are at least 70% homologous to the
CC polypeptides. The method of the invention has cytostatic activity, and
CC may have a use in gene therapy. The method is useful in identifying
CC markers specific for one or several types of cancer, depending on the
CC tissue origin, which may be used in numerous diagnostic and prognostic
CC applications as well as cancer type-specific targets for therapeutic
CC intervention. The compounds that modulate the activity of a tumour
CC suppressor gene are useful in the treatment of cancer or as anti-cancer
CC drugs. The present sequence represents a polynucleotide of the invention.
XX SQ Sequence 4887 BP; 1238 A; 1327 C; 1071 G; 1251 T; 0 U; 0 Other;
Query Match 68.7%; Score 20.6; DB 9; Length 4887;
Best Local Similarity 85.2%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCGAGGAGTAGCAAT 27
DB 1910 TGGCGCTGAACCGAGGAGTAGCAAGAGT 1884
RESULT 14
AAS11651/C
ID AAS11651 standard; DNA; 41 BP.
XX AC AAS11651;
XX DT 24-OCT-2001 (first entry)
XX DE Human CYP2B6 allele variant polynucleotide fragment #6.
XX CYP2B6; cytostatic; gene therapy; genotyping; cancer; metabolism; ds;
XX KW human; cancer susceptibility; environmental carcinogen.
XX OS Homo sapiens.
XX PN WO200159152-A2.
XX XX

PD 16-AUG-2001.
XX 09-FEB-2001; 2001WO-EP001456.
XX 09-FEB-2000; 2000EP-00102701.
XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX Zanger UM, Lang T;
XX WPI; 2001-502719/55.
XX New polynucleotide(s) of the polymorphic human CYP2B6 gene for the
XX detection and treatment of disorders i.e. cancer.
XX Claim 1; Page 80; 83pp; English.
XX The sequence represents a polynucleotide variant sequence of the human
XX CYP2B6 gene. It is used for specific detection and genotyping of CYP2B6
XX alleles in humans, determination of which is useful for the optimisation
XX of therapies utilising CYP2B6 substrates. Oligonucleotide sequences are
XX useful in detection of the individual predisposition to several common
XX cancers caused by environmental carcinogens, and diseases treated with
XX drugs that are targets of the CYP2B6 gene product, whose metabolism is
XX therefore dependent on CYP2B6. Cancer or susceptibility to cancer can be
XX diagnosed by detecting the presence of a molecular variant of CYP2B6.
XX From variants of the alleles, modulators of the activity can be developed
XX for use in treatment and prevention of CYP2B6-related disorders
XX SQ Sequence 41 BP; 8 A; 17 C; 6 G; 10 T; 0 U; 0 Other;
Query Match 68.0%; Score 20.4; DB 4; Length 41;
Best Local Similarity 95.8%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCGAGGAGTAGCA 22
DB 25 TGGCGCTGAACCGAGGAGTAGCA 4
RESULT 15
AAH69361
ID AAH69361 standard; cDNA; 590 BP.
XX AC AAH69361;
XX DT 19-SEP-2001 (first entry)
XX Human cervical cancer marker nucleic acid 635.
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX Homo sapiens.
XX WO200142467-A2.
XX 14-JUN-2001.
XX 08-DEC-2000; 2000WO-US033312.
XX 08-DEC-1999; 99US-0169681P.
XX 21-DEC-1999; 99US-0171350P.
XX 14-MAR-2000; 2000US-0189315P.
XX 12-MAY-2000; 2000US-0203791P.
XX 09-JUN-2000; 2000US-0210600P.
XX 21-JUL-2000; 2000US-0220114P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI; 2001-375006/39.
XX DR 200159152-A2.
XX XX

PT New isolated nucleic acid for diagnosing and treating cervical cancer and
PT for assessing and detecting compounds for treating the cancer.

XX
PS Claim 1; Page 214; 1051pp; English.

XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy

XX
SQ Sequence 590 BP; 227 A; 110 C; 98 G; 153 T; 0 U; 2 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 590;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29
DB 409 CACAGAACCCAGGAGTAGCACCACCAATCA 434

RESULT 16
AAH72227

ID AAH72227 standard; cDNA; 649 BP.

XX
AC AAH72227;

XX
DT 19-SEP-2001 (first entry)

XX
DE Human cervical cancer marker nucleic acid 3501.

XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX
OS Homo sapiens.

XX
PN WO200142467-A2.

XX
PD 14-JUN-2001.

XX
PF 08-DEC-2000; 2000WO-US033312.

XX
PR 08-DEC-1999; 99US-0169681P.

XX
PR 21-DEC-1999; 99US-0171350P.

XX
PR 14-MAR-2000; 2000US-0189315P.

XX
PR 12-MAY-2000; 2000US-0203791P.

XX
PR 09-JUN-2000; 2000US-0210600P.

XX
PR 21-JUL-2000; 2000US-0220114P.

XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX
PI Schlegel R, Deeds J, Berger A, Zhao X;

XX
DR WPI; 2001-375006/39.

XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer and
XX for assessing and detecting compounds for treating the cancer.

XX
PS Claim 1; Page 669; 1051pp; English.

XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy

XX
SQ Sequence 649 BP; 253 A; 118 C; 108 G; 170 T; 0 U; 0 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 649;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29
DB 407 CACAGAACCCAGGAGTAGCACCACCAATCA 432

RESULT 17
AAH71164

ID AAH71164 standard; cDNA; 778 BP.

XX
AC AAH71164;

XX
DT 19-SEP-2001 (first entry)

XX
DE Human cervical cancer marker nucleic acid 2438.

XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX
OS Homo sapiens.

XX
PN WO200142467-A2.

XX
PD 14-JUN-2001.

XX
PF 08-DEC-2000; 2000WO-US033312.

XX
PR 08-DEC-1999; 99US-0169681P.

XX
PR 21-DEC-1999; 99US-0171350P.

XX
PR 14-MAR-2000; 2000US-0189315P.

XX
PR 12-MAY-2000; 2000US-0203791P.

XX
PR 09-JUN-2000; 2000US-0210600P.

XX
PR 21-JUL-2000; 2000US-0220114P.

XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX
PI Schlegel R, Deeds J, Berger A, Zhao X;

XX
DR WPI; 2001-375006/39.

XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer and
XX for assessing and detecting compounds for treating the cancer.

XX
PS Claim 1; Page 511-512; 1051pp; English.

XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy

XX
SQ Sequence 778 BP; 280 A; 153 C; 135 G; 207 T; 0 U; 3 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 778;
Best Local Similarity 84.6%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29
DB 420 CACAGAACCCAGGAGTAGCACCACCAATCA 445

RESULT 18
AAH72859/C

ID AAH72859 standard; cDNA; 2420 BP.

XX
AC AAH72859;

XX 19-SEP-2001 (first entry)
DE Human cervical cancer marker nucleic acid 4133.
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX Homo sapiens.
XX WO200142467-A2.
XX 14-JUN-2001.
XX 08-DEC-2000; 2000WO-US033312.
XX 08-DEC-1999; 99US-0169681P.
XX 21-DEC-1999; 99US-0171350P.
XX 14-MAR-2000; 2000US-0189315P.
XX 13-MAY-2000; 2000US-0203791P.
XX 09-JUN-2000; 2000US-0210600P.
XX 21-JUL-2000; 2000US-0220114P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI; 2001-375006/39.
XX New isolated nucleic acid for diagnosing and treating cervical cancer and
XX for assessing and detecting compounds for treating the cancer.
XX Claim 1; Page 894; 1051pp; English.
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
XX cervical cancer with cytostatic activity. The nucleic acids and encoded
XX polypeptides are useful to assess if a patient is afflicted with
XX cervical cancer or has a pre-malignant condition; to monitor the
XX progression of cervical cancer or a premalignant condition in a patient;
XX and to select and/or assess the efficacy of a compound or therapy for
XX inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy
XX Sequence 2420 BP; 641 A; 458 C; 473 G; 838 T; 0 U; 10 Other;
Query Match 65.3%; Score 19.6; DB 4; Length 2420;
Best Local Similarity 84.6%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 CGCAGACCCAGGAGTACACCATGA 29
DB 1546 CACAGAACCCAGGATAGCACCATCA 1521
RESULT 19
ABV25077/c
ID ABV25077 standard; cDNA; 2420 BP.
XX AC ABV25077;
XX 16-SEP-2002 (first entry)
XX Human prostate expression marker cDNA 25068.
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX Homo sapiens.
XX WO200160860-A2.
XX 23-AUG-2001.
XX 20-FEB-2001; 2001WO-US005171.
XX

XX 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
XX 25-MAY-2000; 2000US-0207454P.
XX 09-JUN-2000; 2000US-0211314P.
XX 18-JUL-2000; 2000US-0219007P.
XX 13-DEC-2000; 2000US-0255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 4873-4874; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 2420 BP; 641 A; 458 C; 473 G; 838 T; 0 U; 10 Other;
Query Match 65.3%; Score 19.6; DB 5; Length 2420;
Best Local Similarity 84.6%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 CGCAGACCCAGGAGTACACCATGA 29
DB 1546 CACAGAACCCAGGATAGCACCATCA 1521
RESULT 20
ABL27011
ID ABL27011 standard; DNA; 2427 BP.
XX AC ABL27011;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 32506.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li FWD, Myers EW;
XX WPI; 2001-656860/75.
XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX
PS Claim 1; SEQ ID NO 32506; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161176-AB161175), expressed DNA
CC sequences (AB161176-AB161175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2427 BP; 681 A; 694 C; 681 G; 371 T; 0 U; 0 Other;
SQ
Query Match 65.3%; Score 19.6; DB 4; Length 2427;
Best Local Similarity 84.6%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 5 GCAGAACCCAGGAGTAGCCCAATGAG 30
Db 2138 GCAGCACCAGGAGCAGCAACAATCAG 2163
RESULT 21
ABL27010
ID ABL27010 standard; DNA; 6200 BP.
XX
AC ABL27010;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 32503.
XX
XX Drosophila; developmental biology; cell signaling; insecticide;
KW pharmaceutical; gene; ds.
KW
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEXE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 32503; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161176-AB161175), expressed DNA
CC sequences (AB161176-AB161175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 6200 BP; 1834 A; 1436 C; 1429 G; 1501 T; 0 U; 0 Other;
SQ
Query Match 65.3%; Score 19.6; DB 4; Length 6200;
Best Local Similarity 84.6%; Pred. No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 5 GCAGAACCCAGGAGTAGCCCAATGAG 30
Db 4454 GCAGCACCAGGAGCAGCAACAATCAG 4479
RESULT 22
ABA61254
ID ABA61254 standard; DNA; 537 BP.
XX
AC ABA61254;
XX
XX 01-FEB-2002 (first entry)
DT
XX Human foetal liver single exon nucleic acid probe #9559.
DE
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
OS
XX WO200157277-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human foetal liver.
XX
XX Claim 1; SEQ ID NO 9559; 639pp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
SQ
Query Match 64.7%; Score 19.4; DB 4; Length 537;
Best Local Similarity 79.3%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 GGCGCAGAACCCAGGAGTAGCCCAATGAG 30
Db 49 GGCCACACACAGCAGAGGCTCCAGTCAG 77

```
RESULT 23
AAI41155
ID AAI41155 standard; DNA; 537 BP.
XX
XX AAI41155;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #9841 used to measure gene expression in human placenta sample.
DE
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 25; SEQ ID NO 9841; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders
XX
XX Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
XX
XX Query Match 64.7%; Score 19.4; DB 4; Length 537;
XX Best Local Similarity 79.3%; Pred. No. 2.7e+02;
XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 2 GGCGCAGAACCCAGGAGTAGCACCACATGAG 30
XX |||||
XX 49 GGCCGAGAACCCAGGAGTAGCACCACATGAG 77
XX

RESULT 24
ABA29093
ID ABA29093 standard; DNA; 537 BP.
XX
XX ABA29093;
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #7559 for gene expression analysis in human heart cell sample.
DE
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
XX Homo sapiens.
XX
```

```
PN WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000666.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 1; SEQ ID NO 7559; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease. Note: The sequence data for this patent did not
XX form part of the printed specification, but was obtained in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
XX
XX Query Match 64.7%; Score 19.4; DB 4; Length 537;
XX Best Local Similarity 79.3%; Pred. No. 2.7e+02;
XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 2 GGCGCAGAACCCAGGAGTAGCACCACATGAG 30
XX |||||
XX 49 GGCCGAGAACCCAGGAGTAGCACCACATGAG 77
XX

RESULT 25
AAK35442
ID AAK35442 standard; DNA; 537 BP.
XX
XX AAK35442;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 9999.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX
```

```
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 9999; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 4; Length 537;
Best Local Similarity 79.3%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
PS Example 4; SEQ ID NO 9999; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 4; Length 537;
Best Local Similarity 79.3%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
PS Example 4; SEQ ID NO 9999; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
Query Match 64.7%; Score 19.4; DB 4; Length 537;
Best Local Similarity 79.3%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
QY 2 GGCGCAGAACCCAGGAGTAGCACAATGAG 30
DB 49 GGCCCGAAGAACCCAGGAGTAGCACAATGAG 77
XX
RESULT 26
AAK09550
ID AAK09550 standard; DNA; 537 BP.
XX
AC AAK09550;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 9541.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 1; SEQ ID NO 10167; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. A825011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
```

CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC
SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 4; Length 537;
Best Local Similarity 79.3%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGCGCAGAACCCAGGAGTAGCACCAGAG 30
Db 49 GGCCGAGAACCCAGGAGGCTCCAGTGAG 77

RESULT 28
ABS09801
ID ABS09801 standard; DNA; 537 BP.
XX
AC ABS09801;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID NO 9792.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN W0200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US0000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 1; SEQ ID NO 9792; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC ; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)

CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a single exon probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;

Query Match 64.7%; Score 19.4; DB 6; Length 537;
Best Local Similarity 79.3%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGCGCAGAACCCAGGAGTAGCACCAGAG 30
Db 49 GGCCGAGAACCCAGGAGGCTCCAGTGAG 77

RESULT 29
AAV52169
ID AAV52169 standard; DNA; 21706 BP.
XX
AC AAV52169;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:36.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
KW Streptococcus pneumoniae.
XX
OS
XX
PN W09818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US019588.
XX
PR 31-OCT-1996; 96US-0029960P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI Dougherty BA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
PS Claim 1; Page 358-371; 1409pp; English.

Job time : 187.506 secs

XX CC The present invention describes a computer readable medium which has the
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC 391, identifying members of the library which contain sequences that
CC hybridize to the target sequence and isolating the nucleic acid molecules
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC homologous to amplification primers derived from the fragment of the S.
CC pneumoniae genome to prime the amplification and isolating the amplified
CC sequences. The computer readable medium can be used in a computer-based
CC system for identifying fragments of the S. pneumoniae genome of
CC commercial importance, or expression modulating fragments of the S.
CC pneumoniae genome. Products from the present invention can be used in
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC for S. pneumoniae

XX SQ Sequence 21706 BP; 6489 A; 3945 C; 4879 G; 6392 T; 0 U; 0 Other;

Query Match 64.0%; Score 19.2; DB 2; Length 21706;
Best Local Similarity 87.5%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCCCAATG 28

Db 6961 GAAGAACCAGAGTAGCTCCAATG 6984

RESULT 30

ABS56454_14/c

Continuation (15 of 22) of ABS56454 from base 1400001 (Streptococcus pneumoniae type 4 s
WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | ABS56454_00 | 1 | 110000 |
| WP | ABS56454_01 | 100001 | 210000 |
| WP | ABS56454_02 | 200001 | 310000 |
| WP | ABS56454_03 | 300001 | 410000 |
| WP | ABS56454_04 | 400001 | 510000 |
| WP | ABS56454_05 | 500001 | 610000 |
| WP | ABS56454_06 | 600001 | 710000 |
| WP | ABS56454_07 | 700001 | 810000 |
| WP | ABS56454_08 | 800001 | 910000 |
| WP | ABS56454_09 | 900001 | 1010000 |
| WP | ABS56454_10 | 1000001 | 1110000 |
| WP | ABS56454_11 | 1100001 | 1210000 |
| WP | ABS56454_12 | 1200001 | 1310000 |
| WP | ABS56454_13 | 1300001 | 1410000 |
| WP | ABS56454_14 | 1400001 | 1510000 |
| WP | ABS56454_15 | 1500001 | 1610000 |
| WP | ABS56454_16 | 1600001 | 1710000 |
| WP | ABS56454_17 | 1700001 | 1810000 |
| WP | ABS56454_18 | 1800001 | 1910000 |
| WP | ABS56454_19 | 1900001 | 2010000 |
| WP | ABS56454_20 | 2000001 | 2110000 |
| WP | ABS56454_21 | 2100001 | 2162598 |

Query Match 64.0%; Score 19.2; DB 7; Length 110000;
Best Local Similarity 87.5%; Pred. No. 7.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCCCAATG 28

Db 56073 GAAGAACCAGAGTAGCTCCAATG 56050

Search completed: June 20, 2004, 10:18:21

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1401.9 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgagaccagtagtagcaacaaatgag 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 21.2 | 70.7 | 670 | 13 BY763514 | BY763514 BX763514 |
| 2 | 21.2 | 70.7 | 691 | 13 CB551691 | CB551691 MMSPP0057 |
| 3 | 20.8 | 69.3 | 298 | 10 BB025068 | BB025068 BB025068 |
| 4 | 20.8 | 69.3 | 839 | 14 CF593587 | CF593587 AGENCOURT |

| | | | | | | | | | |
|----------|-----------|-----|----|----------|-----|------|------|---|----|
| CK130228 | AGENCOURT | 767 | 14 | CK130228 | 767 | 68.7 | 20.6 | c | 5 |
| CK130232 | AGENCOURT | 767 | 14 | CK130232 | 767 | 68.7 | 20.6 | c | 6 |
| CF271960 | AGENCOURT | 768 | 14 | CF271960 | 768 | 68.7 | 20.6 | c | 7 |
| CK130229 | AGENCOURT | 771 | 14 | CK130229 | 771 | 68.7 | 20.6 | c | 8 |
| CK130230 | AGENCOURT | 779 | 14 | CK130230 | 779 | 68.7 | 20.6 | c | 9 |
| CF271962 | AGENCOURT | 795 | 14 | CF271962 | 795 | 68.7 | 20.6 | c | 10 |
| CK130231 | AGENCOURT | 804 | 14 | CK130231 | 804 | 68.7 | 20.6 | c | 11 |
| BZ773303 | mcv67d11. | 986 | 28 | BZ773303 | 986 | 68.7 | 20.6 | c | 12 |
| BF444251 | 262053 MA | 227 | 10 | BF444251 | 227 | 68.0 | 20.4 | c | 13 |
| BQ136744 | NF002D06S | 230 | 13 | BQ136744 | 230 | 68.0 | 20.4 | c | 14 |
| BB246711 | BB246711 | 241 | 10 | BB246711 | 241 | 68.0 | 20.4 | c | 15 |
| BE318013 | NF061H03L | 250 | 10 | BE318013 | 250 | 68.0 | 20.4 | c | 16 |
| BE316071 | NF031B01L | 281 | 10 | BE316071 | 281 | 68.0 | 20.4 | c | 17 |
| BF640484 | NF037A02I | 290 | 10 | BF640484 | 290 | 68.0 | 20.4 | c | 18 |
| BI266129 | NF100G10I | 293 | 12 | BI266129 | 293 | 68.0 | 20.4 | c | 19 |
| BI267471 | NF105E03I | 312 | 12 | BI267471 | 312 | 68.0 | 20.4 | c | 20 |
| BI265945 | NF092F12I | 319 | 12 | BI265945 | 319 | 68.0 | 20.4 | c | 21 |
| AW127696 | M110445 D | 349 | 10 | AW127696 | 349 | 68.0 | 20.4 | c | 22 |
| BE323556 | NF010H11P | 358 | 10 | BE323556 | 358 | 68.0 | 20.4 | c | 23 |
| BE315922 | NF028C04L | 361 | 10 | BE315922 | 361 | 68.0 | 20.4 | c | 24 |
| AW683712 | NF018A04L | 364 | 10 | AW683712 | 364 | 68.0 | 20.4 | c | 25 |
| BE324258 | NF020A04P | 382 | 10 | BE324258 | 382 | 68.0 | 20.4 | c | 26 |
| BE322941 | NF025C04I | 386 | 10 | BE322941 | 386 | 68.0 | 20.4 | c | 27 |
| BE323559 | NF003F04P | 386 | 10 | BE323559 | 386 | 68.0 | 20.4 | c | 28 |
| BE249756 | NF021H07L | 389 | 10 | BE249756 | 389 | 68.0 | 20.4 | c | 29 |
| BI263513 | NF087A01P | 398 | 12 | BI263513 | 398 | 68.0 | 20.4 | c | 30 |
| AW911825 | U88B003.Y | 409 | 10 | AW911825 | 409 | 68.0 | 20.4 | c | 31 |
| BE323980 | NF012A07P | 425 | 10 | BE323980 | 425 | 68.0 | 20.4 | c | 32 |
| BF005193 | EST43369I | 428 | 10 | BF005193 | 428 | 68.0 | 20.4 | c | 33 |
| BE318665 | NF040G06L | 438 | 10 | BE318665 | 438 | 68.0 | 20.4 | c | 34 |
| BI265760 | NF086E06L | 447 | 12 | BI265760 | 447 | 68.0 | 20.4 | c | 35 |
| BE317431 | NF069C12L | 458 | 10 | BE317431 | 458 | 68.0 | 20.4 | c | 36 |
| BF464422 | UI-X-CG0P | 462 | 10 | BF464422 | 462 | 68.0 | 20.4 | c | 37 |
| BE319812 | NF042C05L | 473 | 10 | BE319812 | 473 | 68.0 | 20.4 | c | 38 |
| BE324767 | NF017C05P | 505 | 10 | BE324767 | 505 | 68.0 | 20.4 | c | 39 |
| BE324843 | NF021C06P | 507 | 10 | BE324843 | 507 | 68.0 | 20.4 | c | 40 |
| BF640358 | NF031H09I | 513 | 10 | BF640358 | 513 | 68.0 | 20.4 | c | 41 |
| BF520833 | EST458306 | 523 | 12 | BF520833 | 523 | 68.0 | 20.4 | c | 42 |
| BG448827 | NF002B06I | 523 | 12 | BG448827 | 523 | 68.0 | 20.4 | c | 43 |
| BE316819 | NF066D08L | 526 | 10 | BE316819 | 526 | 68.0 | 20.4 | c | 44 |
| BI264512 | NF108E04P | 528 | 12 | BI264512 | 528 | 68.0 | 20.4 | c | 45 |
| BE318717 | NF042D05L | 539 | 10 | BE318717 | 539 | 68.0 | 20.4 | c | 46 |
| BQ139956 | NF027B08P | 539 | 13 | BQ139956 | 539 | 68.0 | 20.4 | c | 47 |
| BF637160 | NF049H04L | 543 | 10 | BF637160 | 543 | 68.0 | 20.4 | c | 48 |
| BI263324 | NF088H07P | 557 | 12 | BI263324 | 557 | 68.0 | 20.4 | c | 49 |
| BE124199 | EST394324 | 558 | 10 | BE124199 | 558 | 68.0 | 20.4 | c | 50 |
| BF640661 | NF044D10I | 561 | 10 | BF640661 | 561 | 68.0 | 20.4 | c | 51 |
| BE319051 | NF034D11L | 584 | 10 | BE319051 | 584 | 68.0 | 20.4 | c | 52 |
| BF639376 | NF011F04I | 570 | 10 | BF639376 | 570 | 68.0 | 20.4 | c | 53 |
| AW683744 | NF018D07L | 576 | 10 | AW683744 | 576 | 68.0 | 20.4 | c | 54 |
| BF638168 | NF044G01P | 577 | 10 | BF638168 | 577 | 68.0 | 20.4 | c | 55 |
| BF006139 | EST434709 | 577 | 10 | BF006139 | 577 | 68.0 | 20.4 | c | 56 |
| BF004897 | EST433458 | 581 | 10 | BF004897 | 581 | 68.0 | 20.4 | c | 57 |
| BE316701 | NF065A06L | 582 | 10 | BE316701 | 582 | 68.0 | 20.4 | c | 58 |
| BF639007 | NF094B12P | 583 | 10 | BF639007 | 583 | 68.0 | 20.4 | c | 59 |
| BF637828 | NF043D05P | 587 | 10 | BF637828 | 587 | 68.0 | 20.4 | c | 60 |
| AW692465 | NF051G09S | 590 | 10 | AW692465 | 590 | 68.0 | 20.4 | c | 61 |
| BE323679 | NF005A11P | 595 | 10 | BE323679 | 595 | 68.0 | 20.4 | c | 62 |
| BF005033 | EST433531 | 597 | 10 | BF005033 | 597 | 68.0 | 20.4 | c | 63 |
| AW980360 | EST391513 | 602 | 10 | AW980360 | 602 | 68.0 | 20.4 | c | 64 |
| BE324591 | NF014E04P | 604 | 10 | BE324591 | 604 | 68.0 | 20.4 | c | 65 |
| BI268290 | NF120D11I | 605 | 12 | BI268290 | 605 | 68.0 | 20.4 | c | 66 |
| BF642786 | NF072F04I | 608 | 10 | BF642786 | 608 | 68.0 | 20.4 | c | 67 |
| BG457889 | NF033B03P | 617 | 12 | BG457889 | 617 | 68.0 | 20.4 | c | 68 |
| AW683772 | NF018G11L | 621 | 10 | AW683772 | 621 | 68.0 | 20.4 | c | 69 |
| BE324010 | NF012D11P | 623 | 10 | BE324010 | 623 | 68.0 | 20.4 | c | 70 |
| BQ138571 | NF004G07P | 630 | 13 | BQ138571 | 630 | 68.0 | 20.4 | c | 71 |
| AW256790 | EST304927 | 631 | 10 | AW256790 | 631 | 68.0 | 20.4 | c | 72 |
| BG456877 | NF088F10P | 633 | 12 | BG456877 | 633 | 68.0 | 20.4 | c | 73 |
| BM914057 | EST592150 | 634 | 12 | BM914057 | 634 | 68.0 | 20.4 | c | 74 |
| BQ138282 | NF001E06P | 635 | 13 | BQ138282 | 635 | 68.0 | 20.4 | c | 75 |
| BG456673 | NF092E07P | 639 | 12 | BG456673 | 639 | 68.0 | 20.4 | c | 76 |
| BF520572 | EST458045 | 640 | 10 | BF520572 | 640 | 68.0 | 20.4 | c | 77 |

| | | | | | | | | | |
|-------|-----|----|----------|-------|------|------|------|----|----------|
| C 78 | 644 | 12 | BQ456201 | 151 | 19.4 | 64.7 | 289 | 13 | BQ305306 |
| C 79 | 645 | 10 | BE317057 | C 152 | 19.4 | 64.7 | 291 | 9 | AA344653 |
| C 80 | 646 | 13 | BQ138629 | C 153 | 19.4 | 64.7 | 305 | 10 | BE655367 |
| C 81 | 647 | 10 | BQ249506 | C 154 | 19.4 | 64.7 | 309 | 12 | BE559599 |
| C 82 | 648 | 10 | BQ637864 | C 155 | 19.4 | 64.7 | 328 | 29 | CG707991 |
| C 83 | 649 | 10 | BQ637864 | C 156 | 19.4 | 64.7 | 330 | 9 | AI426118 |
| C 84 | 650 | 10 | BQ637864 | C 157 | 19.4 | 64.7 | 332 | 13 | BY173156 |
| C 85 | 651 | 10 | BQ637864 | C 158 | 19.4 | 64.7 | 336 | 29 | CG707847 |
| C 86 | 652 | 10 | BQ637864 | C 159 | 19.4 | 64.7 | 338 | 9 | AA188722 |
| C 87 | 653 | 10 | BQ637864 | C 160 | 19.4 | 64.7 | 339 | 14 | W82305 |
| C 88 | 654 | 10 | BQ637864 | C 161 | 19.4 | 64.7 | 411 | 9 | AI182795 |
| C 89 | 655 | 10 | BQ637864 | C 162 | 19.4 | 64.7 | 422 | 13 | BY158314 |
| C 90 | 656 | 10 | BQ637864 | C 163 | 19.4 | 64.7 | 426 | 13 | BY542403 |
| C 91 | 657 | 10 | BQ637864 | C 164 | 19.4 | 64.7 | 429 | 9 | AA968188 |
| C 92 | 658 | 10 | BQ637864 | C 165 | 19.4 | 64.7 | 430 | 13 | BY551293 |
| C 93 | 659 | 10 | BQ637864 | C 166 | 19.4 | 64.7 | 432 | 14 | CB820248 |
| C 94 | 660 | 10 | BQ637864 | C 167 | 19.4 | 64.7 | 437 | 9 | AA671087 |
| C 95 | 661 | 10 | BQ637864 | C 168 | 19.4 | 64.7 | 438 | 10 | BB825444 |
| C 96 | 662 | 10 | BQ637864 | C 169 | 19.4 | 64.7 | 444 | 9 | AA656665 |
| C 97 | 663 | 10 | BQ637864 | C 170 | 19.4 | 64.7 | 446 | 13 | BY545446 |
| C 98 | 664 | 10 | BQ637864 | C 171 | 19.4 | 64.7 | 452 | 13 | BY552318 |
| C 99 | 665 | 10 | BQ637864 | C 172 | 19.4 | 64.7 | 453 | 10 | AW908257 |
| C 100 | 666 | 10 | BQ637864 | C 173 | 19.4 | 64.7 | 458 | 28 | AQ335940 |
| C 101 | 667 | 10 | BQ637864 | C 174 | 19.4 | 64.7 | 463 | 13 | BQ297619 |
| C 102 | 668 | 10 | BQ637864 | C 175 | 19.4 | 64.7 | 465 | 10 | BB653543 |
| C 103 | 669 | 10 | BQ637864 | C 176 | 19.4 | 64.7 | 465 | 10 | BB822361 |
| C 104 | 670 | 10 | BQ637864 | C 177 | 19.4 | 64.7 | 468 | 13 | BY598882 |
| C 105 | 671 | 10 | BQ637864 | C 178 | 19.4 | 64.7 | 479 | 13 | BY448067 |
| C 106 | 672 | 10 | BQ637864 | C 179 | 19.4 | 64.7 | 479 | 13 | BY689321 |
| C 107 | 673 | 10 | BQ637864 | C 180 | 19.4 | 64.7 | 488 | 14 | N74643 |
| C 108 | 674 | 10 | BQ637864 | C 181 | 19.4 | 64.7 | 500 | 28 | AZ179652 |
| C 109 | 675 | 10 | BQ637864 | C 182 | 19.4 | 64.7 | 503 | 14 | CB235083 |
| C 110 | 676 | 10 | BQ637864 | C 183 | 19.4 | 64.7 | 513 | 28 | BD230131 |
| C 111 | 677 | 10 | BQ637864 | C 184 | 19.4 | 64.7 | 519 | 29 | CE395685 |
| C 112 | 678 | 10 | BQ637864 | C 185 | 19.4 | 64.7 | 531 | 10 | BB772109 |
| C 113 | 679 | 10 | BQ637864 | C 186 | 19.4 | 64.7 | 536 | 29 | CG726707 |
| C 114 | 680 | 10 | BQ637864 | C 187 | 19.4 | 64.7 | 566 | 28 | AZ747080 |
| C 115 | 681 | 10 | BQ637864 | C 188 | 19.4 | 64.7 | 583 | 28 | AZ758572 |
| C 116 | 682 | 10 | BQ637864 | C 189 | 19.4 | 64.7 | 585 | 10 | BE402590 |
| C 117 | 683 | 10 | BQ637864 | C 190 | 19.4 | 64.7 | 585 | 13 | BQ508170 |
| C 118 | 684 | 10 | BQ637864 | C 191 | 19.4 | 64.7 | 601 | 9 | AI981752 |
| C 119 | 685 | 10 | BQ637864 | C 192 | 19.4 | 64.7 | 633 | 14 | CD776569 |
| C 120 | 686 | 10 | BQ637864 | C 193 | 19.4 | 64.7 | 634 | 10 | BE368075 |
| C 121 | 687 | 10 | BQ637864 | C 194 | 19.4 | 64.7 | 672 | 12 | BI295028 |
| C 122 | 688 | 10 | BQ637864 | C 195 | 19.4 | 64.7 | 672 | 12 | BQ321522 |
| C 123 | 689 | 10 | BQ637864 | C 196 | 19.4 | 64.7 | 678 | 10 | BB506736 |
| C 124 | 690 | 10 | BQ637864 | C 197 | 19.4 | 64.7 | 683 | 28 | BH046494 |
| C 125 | 691 | 10 | BQ637864 | C 198 | 19.4 | 64.7 | 693 | 29 | CC685799 |
| C 126 | 692 | 10 | BQ637864 | C 199 | 19.4 | 64.7 | 699 | 12 | BG974597 |
| C 127 | 693 | 10 | BQ637864 | C 200 | 19.4 | 64.7 | 702 | 29 | CC621439 |
| C 128 | 694 | 10 | BQ637864 | C 201 | 19.4 | 64.7 | 702 | 29 | CC621445 |
| C 129 | 695 | 10 | BQ637864 | C 202 | 19.4 | 64.7 | 704 | 29 | CC675967 |
| C 130 | 696 | 10 | BQ637864 | C 203 | 19.4 | 64.7 | 711 | 29 | CC688083 |
| C 131 | 697 | 10 | BQ637864 | C 204 | 19.4 | 64.7 | 711 | 29 | CC688083 |
| C 132 | 698 | 10 | BQ637864 | C 205 | 19.4 | 64.7 | 734 | 28 | BZ526088 |
| C 133 | 699 | 10 | BQ637864 | C 206 | 19.4 | 64.7 | 734 | 28 | BZ526088 |
| C 134 | 700 | 10 | BQ637864 | C 207 | 19.4 | 64.7 | 796 | 10 | BE034896 |
| C 135 | 701 | 10 | BQ637864 | C 208 | 19.4 | 64.7 | 804 | 12 | BG329720 |
| C 136 | 702 | 10 | BQ637864 | C 209 | 19.4 | 64.7 | 805 | 29 | CC558304 |
| C 137 | 703 | 10 | BQ637864 | C 210 | 19.4 | 64.7 | 808 | 14 | CB819759 |
| C 138 | 704 | 10 | BQ637864 | C 211 | 19.4 | 64.7 | 821 | 29 | CC610468 |
| C 139 | 705 | 10 | BQ637864 | C 212 | 19.4 | 64.7 | 838 | 29 | CG345029 |
| C 140 | 706 | 10 | BQ637864 | C 213 | 19.4 | 64.7 | 839 | 29 | CG325207 |
| C 141 | 707 | 10 | BQ637864 | C 214 | 19.4 | 64.7 | 841 | 29 | CC604958 |
| C 142 | 708 | 10 | BQ637864 | C 215 | 19.4 | 64.7 | 849 | 28 | BH046796 |
| C 143 | 709 | 10 | BQ637864 | C 216 | 19.4 | 64.7 | 893 | 29 | CC604952 |
| C 144 | 710 | 10 | BQ637864 | C 217 | 19.4 | 64.7 | 930 | 14 | CF215604 |
| C 145 | 711 | 10 | BQ637864 | C 218 | 19.4 | 64.7 | 940 | 14 | CB993468 |
| C 146 | 712 | 10 | BQ637864 | C 219 | 19.4 | 64.7 | 964 | 29 | CG325217 |
| C 147 | 713 | 10 | BQ637864 | C 220 | 19.4 | 64.7 | 1111 | 13 | BUI15986 |
| C 148 | 714 | 10 | BQ637864 | C 221 | 19.4 | 64.7 | 1265 | 11 | AK052797 |
| C 149 | 715 | 10 | BQ637864 | C 222 | 19.4 | 64.7 | 1265 | 11 | AK052797 |
| C 150 | 716 | 10 | BQ637864 | C 223 | 19.4 | 64.7 | 1633 | 11 | AK015988 |

| | | | | | | | | |
|--|------|------|------|----|----------|------------|----------|----------|
| 224 | 19.4 | 64.7 | 1895 | 10 | BF124725 | 601761404 | CB499546 | ssalrkc0 |
| C 225 | 19.4 | 64.7 | 1992 | 11 | AK030007 | Mus muscu | 14 | CB499546 |
| 226 | 19.4 | 64.7 | 2224 | 11 | AK031554 | Mus muscu | 29 | CC658109 |
| 227 | 19.4 | 64.7 | 2251 | 11 | BC026385 | Mus muscu | 29 | CC654008 |
| 228 | 19.2 | 64.0 | 287 | 10 | BC366759 | BB366759 | 28 | BZ634921 |
| C 229 | 19.2 | 64.0 | 317 | 13 | BY169754 | BY169754 | 28 | BZ634921 |
| C 230 | 19.2 | 64.0 | 327 | 9 | AI894118 | md82a07.x | | |
| C 231 | 19.2 | 64.0 | 349 | 13 | BY315617 | BY315617 | | |
| 232 | 19.2 | 64.0 | 353 | 14 | W62846 | md82a07.r1 | | |
| 233 | 19.2 | 64.0 | 354 | 9 | AI893206 | md82a07.y | | |
| 234 | 19.2 | 64.0 | 359 | 14 | CF648982 | 3530.1.61 | | |
| 235 | 19.2 | 64.0 | 383 | 13 | BY665821 | BY665821 | | |
| C 236 | 19.2 | 64.0 | 387 | 13 | BX631560 | BX631560 | | |
| C 237 | 19.2 | 64.0 | 391 | 13 | BY165938 | BY165938 | | |
| 238 | 19.2 | 64.0 | 395 | 10 | BF178404 | 601807834 | | |
| C 239 | 19.2 | 64.0 | 402 | 9 | AU008565 | AU008565 | | |
| C 240 | 19.2 | 64.0 | 402 | 13 | EG063795 | H3012H05- | | |
| C 241 | 19.2 | 64.0 | 402 | 13 | BY161437 | BY161437 | | |
| 242 | 19.2 | 64.0 | 403 | 13 | BY505470 | BY505470 | | |
| 243 | 19.2 | 64.0 | 405 | 13 | BY602394 | BY602394 | | |
| C 244 | 19.2 | 64.0 | 409 | 9 | AI849795 | AI849795 | | |
| 245 | 19.2 | 64.0 | 418 | 13 | BY503613 | BY503613 | | |
| C 246 | 19.2 | 64.0 | 423 | 14 | CA896676 | PCEP04269 | | |
| C 247 | 19.2 | 64.0 | 426 | 10 | BF462404 | UI-M-CGOp | | |
| 248 | 19.2 | 64.0 | 445 | 29 | CD242886 | CGWJV05TH | | |
| C 249 | 19.2 | 64.0 | 460 | 14 | CD296898 | StrPu691. | | |
| C 250 | 19.2 | 64.0 | 463 | 9 | AA960429 | ub58b11.s | | |
| 251 | 19.2 | 64.0 | 464 | 9 | AU006636 | AU006636 | | |
| 252 | 19.2 | 64.0 | 473 | 10 | BF454474 | maa03d07. | | |
| 253 | 19.2 | 64.0 | 479 | 9 | AA655629 | vs46h04.x | | |
| 254 | 19.2 | 64.0 | 503 | 9 | AU010308 | AU010308 | | |
| C 255 | 19.2 | 64.0 | 508 | 9 | AU042263 | AU042263 | | |
| C 256 | 19.2 | 64.0 | 508 | 14 | CF646608 | 3530.1.30 | | |
| 257 | 19.2 | 64.0 | 509 | 13 | CA091733 | SCUTM208 | | |
| 258 | 19.2 | 64.0 | 512 | 9 | AI189906 | AUI189906 | | |
| C 259 | 19.2 | 64.0 | 513 | 10 | AW539555 | CO12IC05- | | |
| C 260 | 19.2 | 64.0 | 528 | 12 | BG977705 | ic16h04.x | | |
| C 261 | 19.2 | 64.0 | 530 | 9 | AA265339 | mo82h02.x | | |
| C 262 | 19.2 | 64.0 | 531 | 9 | AU020283 | AU020283 | | |
| C 263 | 19.2 | 64.0 | 532 | 9 | AI988704 | sd06h08.y | | |
| 264 | 19.2 | 64.0 | 537 | 10 | BF454388 | maa04d03. | | |
| C 265 | 19.2 | 64.0 | 543 | 14 | CK137823 | ma97g09. | | |
| C 266 | 19.2 | 64.0 | 547 | 14 | CD296016 | StrPu691. | | |
| C 267 | 19.2 | 64.0 | 559 | 14 | CB572024 | AGEENCOURT | | |
| C 268 | 19.2 | 64.0 | 559 | 10 | BE609902 | 8Q46F04.y | | |
| C 269 | 19.2 | 64.0 | 561 | 10 | AW472125 | si20c03.y | | |
| C 270 | 19.2 | 64.0 | 563 | 9 | AI607786 | ub58b11.x | | |
| C 271 | 19.2 | 64.0 | 563 | 14 | CA137757 | SCQST103 | | |
| C 272 | 19.2 | 64.0 | 565 | 13 | CA137321 | SCAGRT1203 | | |
| C 273 | 19.2 | 64.0 | 570 | 13 | BU927190 | sas96f08. | | |
| C 274 | 19.2 | 64.0 | 574 | 12 | BM939980 | UI-M-CGOp | | |
| C 275 | 19.2 | 64.0 | 575 | 29 | CG346563 | CGVDF13TV | | |
| C 276 | 19.2 | 64.0 | 580 | 28 | AQ330654 | nbxb0047p | | |
| C 277 | 19.2 | 64.0 | 601 | 13 | BQ101895 | 1b7e11.x | | |
| C 278 | 19.2 | 64.0 | 616 | 29 | CG678327 | CG2AB28TV | | |
| 279 | 19.2 | 64.0 | 618 | 13 | BUS44048 | GM880001B | | |
| C 280 | 19.2 | 64.0 | 620 | 14 | CA204616 | SCSGFL1107 | | |
| C 281 | 19.2 | 64.0 | 626 | 12 | BM722053 | UI-E-E0S- | | |
| C 282 | 19.2 | 64.0 | 631 | 28 | CC339003 | CGOAX81TV | | |
| 283 | 19.2 | 64.0 | 637 | 29 | CG213678 | CGVDO49TH | | |
| C 284 | 19.2 | 64.0 | 642 | 14 | CD296222 | StrPu691. | | |
| C 285 | 19.2 | 64.0 | 650 | 29 | CC703220 | CGWHR11TH | | |
| C 286 | 19.2 | 64.0 | 654 | 14 | CA900839 | PCSD13242 | | |
| C 287 | 19.2 | 64.0 | 678 | 14 | CA900841 | PCSC20328 | | |
| C 288 | 19.2 | 64.0 | 685 | 10 | BE632164 | uu12f05.x | | |
| C 289 | 19.2 | 64.0 | 685 | 28 | BZ376889 | 1e75c02.g | | |
| C 290 | 19.2 | 64.0 | 688 | 14 | CF806530 | pHB013XE | | |
| C 291 | 19.2 | 64.0 | 696 | 10 | BE688315 | 1w45d04.x | | |
| C 292 | 19.2 | 64.0 | 700 | 29 | CG057100 | PUING47TD | | |
| C 293 | 19.2 | 64.0 | 711 | 14 | CA896675 | PCEP03767 | | |
| C 294 | 19.2 | 64.0 | 737 | 29 | CC644920 | CGWCB80TV | | |
| C 295 | 19.2 | 64.0 | 738 | 29 | CG239827 | CGVCI161TH | | |
| C 296 | 19.2 | 64.0 | 738 | 29 | CG239827 | CGVCI161TH | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.0 750 29 CC658109</div> <div>19:2 64.0 757 29 CC654008</div> <div>19:2 64.0 761 28 BZ634921</div> <div>19:2 64.0 761 28 BZ634921</div> | | | | | | | | |
| <div>19:2 64.0 750 14 CB499546</div> <div>19:2 64.</div> | | | | | | | | |

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

Location/Qualifiers
1. .670

FEATURES source

/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="taxon:10090"
/clone="E430037109"
/tissue_type="thymus"
/cell_type="thymic cells"
/clone_lib="RIKEN full-length enriched, 2 days neonate thymus thymic cells (NOD)"

ORIGIN

Query Match 70.7%; Score 21.2; DB 13; Length 670;
Best Local Similarity 88.5%; Pred. No. 2.7e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCGCAGAACCCAGGAGTAGCACCAC 26
|||||
Db 108 TGCGCAGAACCCAGGAGTAGCACCAC 133

RESULT 2.

CB551691/c 691 bp mRNA linear EST 01-JUN-2003
LOCUS MWSPO057 H06 WMSP Macaca mulatta cDNA, mRNA sequence.
DEFINITION CB551691
ACCESSION CB551691
VERSION CB551691.1 GI:31300886
KEYWORDS EST.

SOURCE Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.

1 (bases 1 to 691)
Katze, M.C., Bungarner, R., Korth, M., Feldman, R., Amjadi, M. and
Holzman, I.

Expressed sequence tags from Rhesus macaque spleen
Unpublished (2002)
Contact: Holzman T
Katze Lab
University of Washington
Box 358070, Seattle, WA 98195-8070, USA
Tel: 206 732 6156
Fax: 206 732 6055

REFERENCE AUTHORS
Katze, M.C., Bungarner, R., Korth, M., Feldman, R., Amjadi, M. and
Holzman, I.

TITLE Expressed sequence tags from Rhesus macaque spleen

JOURNAL Unpublished (2002)

COMMENT Contact: Holzman T

Katze Lab

University of Washington

Box 358070, Seattle, WA 98195-8070, USA

Tel: 206 732 6156

Fax: 206 732 6055

Email: ted@locke.hs.washington.edu

Similar to GenBank entry AK091287 AK091287 Homo sapiens cDNA

FLJ33968 fls, clone DFNES2001096. 7/2002

Plate: MWSPO057 row: H column: 06.

Location/Qualifiers

1. .691

FEATURES source

/organism="Macaca mulatta"

/mol_type="mRNA"

/db_xref="taxon:9544"
/sex="male"
/cell_type="monocytes"
/dev_stage="adult"
/clone_lib="WMSP"
/note="Organ: spleen"

ORIGIN

Query Match 70.7%; Score 21.2; DB 14; Length 691;
Best Local Similarity 88.5%; Pred. No. 2.7e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29
|||||
Db 424 CGCAGAACCCAGGAGTAGCACCACCAATGA 399

RESULT 3

BB025068

LOCUS

DEFINITION

BB025068 298 bp mRNA linear EST 23-JUN-2000
BB025068 RIKEN full-length enriched, adult male pituitary gland Mus
musculus cDNA clone 5330426a17.3, similar to AJ222586 Mus musculus
mRNA for NEFA protein, mRNA sequence.

ACCESSION BB025068.1 GI:8199385

VERSION BB025068

KEYWORDS Mus musculus (house mouse)

SOURCE EST.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 298)

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,

Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,

Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,

Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K.,

Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,

Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A.,

Watanabe, S., Yamamura, T., Yamahata, I., Yano, R., Yasunishi, A.,

Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and

Hayashizaki, Y.

RIKEN Mouse ESTs (Kono, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Sasaki, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,

Sasai, P., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Iizawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,

Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

Location/Qualifiers

1. .298

FEATURES source

/organism="Mus musculus"

| | | | | |
|--|--|---|--|--|
| <p>/mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="5330426A17"</p> <p>/sex="male"</p> <p>/tissue_type="pituitary gland"</p> <p>/dev_stage="adult"</p> <p>/lab_host="DH10B"</p> <p>/clone_lib="RIKEN full-length enriched, adult male pituitary gland"</p> <p>/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTWN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 3.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAATAATTAATCCCTCCCTCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."</p> | <p>Query Match 69.3%; Score 20.8; DB 10; Length 298;</p> <p>Best Local Similarity 91.7%; Pred. No. 2.9e+03;</p> <p>Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p> | <p>QY 2 GGCGAGAACCCAGGAGTAGCACCA 25</p> <p> </p> <p>DB 49 GGCGAGAACCCAGGAGTAGCATCA 72</p> | <p>ORIGIN</p> <p>LOCUS CF593587 839 bp mRNA linear EST 26-SEP-2003</p> <p>DEFINITION AGENCOURT_15621221 NIH_MGC_147 Homo sapiens cDNA clone IMAGE:30531733 5', mRNA sequence.</p> <p>ACCESSION CF593587</p> <p>VERSION CF593587.1 GI:36347284</p> <p>KEYWORDS EST.</p> <p>SOURCE Homo sapiens (human)</p> <p>ORGANISM Homo sapiens</p> <p>REFERENCE 1 (bases 1 to 839)</p> <p>AUTHORS NIH-MGC http://mgc.nci.nih.gov/.</p> <p>TITLE National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)</p> <p>COMMENT Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov</p> | <p>FEATURES</p> <p>source</p> <p>1..839</p> <p>Location/Qualifiers</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:30531733"</p> <p>/tissue_type="pituitary gland"</p> <p>/dev_stage="adult"</p> <p>/lab_host="DH10B"</p> <p>/clone_lib="RIKEN full-length enriched, adult male pituitary gland"</p> <p>/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTWN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 3.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAATAATTAATCCCTCCCTCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."</p> |
| <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:30531733"</p> <p>/tissue_type="Human Placenta"</p> <p>/lab_host="DH10B Tona"</p> <p>/clone_lib="NIH_MGC_147"</p> <p>/note="Organ: placenta; Vector: pBluescriptR; Site 1: XhoI; Site 2: BamHI; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTWN-3', size-selected for average insert size 2.3 kb and normalized to Rot 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: This is a NIH_MGC library."</p> | <p>Query Match 69.3%; Score 20.8; DB 14; Length 839;</p> <p>Best Local Similarity 91.7%; Pred. No. 4e+03;</p> <p>Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p> | <p>QY 3 GCGCAGAACCCAGGAGTAGCACCA 26</p> <p> </p> <p>DB 18 GCGTAGAACCCAGGAGTAGCACCA 41</p> | <p>ORIGIN</p> <p>LOCUS CK130228 767 bp mRNA linear EST 02-DEC-2003</p> <p>DEFINITION AGENCOURT_15196935 NIH_MGC_195 Homo sapiens cDNA clone IMAGE:7002143 5', mRNA sequence.</p> <p>ACCESSION CK130228</p> <p>VERSION CK130228.1 GI:38622164</p> <p>KEYWORDS EST.</p> <p>SOURCE Homo sapiens (human)</p> <p>ORGANISM Homo sapiens</p> <p>REFERENCE 1 (bases 1 to 767)</p> <p>AUTHORS NIH-MGC http://mgc.nci.nih.gov/.</p> <p>TITLE National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)</p> <p>COMMENT Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-remail.nih.gov</p> | <p>FEATURES</p> <p>source</p> <p>1..767</p> <p>Location/Qualifiers</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:7002143"</p> <p>/tissue_type="mixed"</p> <p>/lab_host="DH5A (T1 phage-resistant)"</p> <p>/clone_lib="NIH_MGC_195"</p> <p>/note="Vector: pDNK-R Dual; Site 1: loxP-Sali; Site 2: loxP-HindIII; Clones from this library have been PCR-amplified using gene-specific primers to contain the complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxP sites</p> |

of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at
ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
a Note: this is a NIH_MGC Library."

ORIGIN

Query Match 68.7%; Score 20.6; DB 14; Length 767;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 72 TGGCGCTGAACCCAGGAGTAGCAGAGCT 46
|||||

RESULT 6

CK130232/c
LOCUS CK130232 767 bp mRNA linear EST 02-DEC-2003
DEFINITION AGENCOURT 15196775 NIH_MGC.195 Homo sapiens cDNA clone
IMAGE:7002138 5', mRNA sequence.

ACCESSION CK130232.1 GI:38622168

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 767)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Narayan Bhat

cDNA Library Preparation: Bhat Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: IRBK4 row: b column: 04

High quality sequence start: 14

High quality sequence stop: 464.

Location/Qualifiers

FEATURES

source

1..767
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7002138"
/tissue_type="mixed"
/lab_host="DH5A (TI phage-resistant)"
/clone_lib="NIH MGC 195"
/note="Vector: pDNR-Dual; Site 1: loxp-Sall; Site 2: loxp-HindIII; Clones from this library have been PCR-amplified using gene-specific primers to contain the complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxp sites of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at
ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
a Note: this is a NIH_MGC Library."

visit our anonymous ftp site at
ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
a Note: this is a NIH_MGC Library."

ORIGIN

Query Match 68.7%; Score 20.6; DB 14; Length 767;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
|||||
Db 107 TGGCGCTGAACCCAGGAGTAGCAGAGCT 81
|||||

RESULT 7

CF271960/c

LOCUS CF271960 768 bp mRNA linear EST 26-NOV-2003

DEFINITION AGENCOURT 15196903 NIH_MGC.195 Homo sapiens cDNA clone

IMAGE:7002142 5', mRNA sequence.

ACCESSION CF271960

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 768)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT On Aug 12, 2003 this sequence version replaced gi:33627872.

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Narayan Bhat

cDNA Library Preparation: Bhat Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: IRBK4 row: b column: 08

High quality sequence stop: 702.

Location/Qualifiers

FEATURES

source

1..768
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7002142"
/tissue_type="mixed"
/lab_host="DH5A (TI phage-resistant)"
/clone_lib="NIH MGC 195"
/note="Vector: pDNR-Dual; Site 1: loxp-Sall; Site 2: loxp-HindIII; Clones from this library have been PCR-amplified using gene-specific primers to contain the complete open reading frame (based on known gene sequences available from NCBI's RefSeq). Template for PCR is cDNA derived from either pooled cytoplasmic polyA RNA from 30 cells lines or pooled total RNA from 10 different tissues (from BD Biosciences/Clontech and Washington University). PCR products are directionally cloned into the loxp sites of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at
ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
a Note: this is a NIH_MGC Library."

ORIGIN

76 TGCGCTGACACGAGGTAGCAAGAT 50

RESULT 9
CK130230/c

LOCUS
CK130230.c

DEFINITION
AGENCOURT_15196839 NIH_MGC_195 Homo sapiens cDNA clone
IMAGE:7002140 5', mRNA sequence.

ACCESSION
CK130230

VERSION
CK130230.1 GI:38622166

KEYWORDS
EST.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS
1 (bases 1 to 779)

TITLE
NIH-MGC <http://imgc.nci.nih.gov/>

JOURNAL
National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: gsabbs@mail.nih.gov
Tissue Procurement: Narayan Bhat
cDNA Library Preparation: Bhat Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: IRBK4 row: b column: 06
High quality sequence stop: 716.
Location/Qualifiers

FEATURES
source

1..779
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7002140"
/tissue_type="mixed"
/lab_host="DH5A (T1 phage-resistant)"
/clone_lib="NIH MGC 195"
/notes="Vector: pDNR-Dual; Site 1: loxp-Sall; Site 2:
loxP-HindIII; Clones from this library have been
PCR-amplified using gene-specific primers to contain the
complete open reading frame (based on known gene sequences
available from NCBI's RefSeq). Template for PCR is cDNA
derived from either pooled cytoplasmic polyA RNA from 30
cells lines or pooled total RNA from 10 different tissues
(from BD Biosciences/Clontech and Washington University).
PCR products are directionally cloned into the loxp sites
of the pDNR-Dual vector. Library constructed by Dr.
Narayan Bhat, Earl Bere III and Hongling Liao (Gene
Expression Laboratory, Research Technology Program, SAIC
Frederick, NCI-Frederick, Frederick, MD 21702). For
information on which gene each clone represents, please
visit our anonymous ftp site at
ftp://image.llnl.gov/image/rearrayed_plates/IRBK4.presv.dat
a Note: this is a NIH_MGC Library."

ORIGIN

Query Match 68.7%; Score 20.6; DB 14; Length 779;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

DEFINITION  AGENCOURT 15196743 NIH MGC 195 Homo sapiens cDNA clone
IMAGE:7002137 5', mRNA sequence.
ACCESSION  CF271962
VERSION    CF271962.2 GI:38559851
KEYWORDS
SOURCE
ORGANISM   Homo sapiens (human)

REFERENCE  1 (bases 1 to 785)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   On Aug 12, 2003 this sequence version replaced gi:33627874.
          Contact: Daniela S. Gerhard, Ph.D.
          Office of Cancer Genomics
          National Cancer Institute / NIH
          Bldg 31 Rm10A07 Bethesda, MD 20892
          Email: cgaabs-r@mail.nih.gov
          Tissue Procurement: Narayan Bhat
          cDNA Library Preparation: Bhat Laboratory
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: IRBK4 row: b column: 03
          High quality sequence start: 4
          High quality sequence stop: 764.

FEATURES             source
     source
1..785
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:7002137"
     /tissue_type="mixed"
     /lab_host="DH5A (T1 phage-resistant)"
     /clone_lib="NIH MGC 195"
     /note="Vector: pDNR-Dual; Site 1: loxP-SalI; Site 2:
     loxP-HindIII; Clones from this library have been
     PCR-amplified using gene-specific primers to contain the
     complete open reading frame (based on known gene sequences
     available from NCBI's RefSeq). Template for PCR is cDNA
     derived from either pooled cytoplasmic polyA RNA from 30
     cells lines or pooled total RNA from 10 different tissues
     (from BD Biosciences/Clontech and Washington University).
     PCR products are directionally cloned into the loxP sites
     of the pDNR-Dual vector. Library constructed by Dr.
     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
     Expression Laboratory, Research Technology Program, SAIC
     Frederick, NCI-Frederick, Frederick, MD 21702). For
     information on which gene each clone represents, please
     visit our anonymous ftp site at
     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
     a Note: this is a NIH_MGC Library."

ORIGIN
Query Match      68.7%; Score 20.6; DB 14; Length 785;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCGAGGTAGCACCAGT 27
    |||||
DB 98 TGGCGCTGAACCGAGGTAGCAGAGT 72
    |||||

RESULT 11
CK130231/c
LOCUS
DEFINITION  AGENCOURT_15196807 NIH MGC 195 Homo sapiens cDNA clone
IMAGE:7002139 5', mRNA sequence.
ACCESSION  CK130231
VERSION    CK130231.1 GI:38622167

```

```

KEYWORDS  EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
           Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 804)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
           Office of Cancer Genomics
           National Cancer Institute / NIH
           Bldg 31 Rm10A07 Bethesda, MD 20892
           Email: cgaabs-r@mail.nih.gov
           Tissue Procurement: Narayan Bhat
           cDNA Library Preparation: Bhat Laboratory
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: IRBK4 row: b column: 05
           High quality sequence start: 3
           High quality sequence stop: 737.

FEATURES             source
     source
1..804
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:7002139"
     /tissue_type="mixed"
     /lab_host="DH5A (T1 phage-resistant)"
     /clone_lib="NIH MGC 195"
     /note="Vector: pDNR-Dual; Site 1: loxP-SalI; Site 2:
     loxP-HindIII; Clones from this library have been
     PCR-amplified using gene-specific primers to contain the
     complete open reading frame (based on known gene sequences
     available from NCBI's RefSeq). Template for PCR is cDNA
     derived from either pooled cytoplasmic polyA RNA from 30
     cells lines or pooled total RNA from 10 different tissues
     (from BD Biosciences/Clontech and Washington University).
     PCR products are directionally cloned into the loxP sites
     of the pDNR-Dual vector. Library constructed by Dr.
     Narayan Bhat, Earl Bere III and Hongling Liao (Gene
     Expression Laboratory, Research Technology Program, SAIC
     Frederick, NCI-Frederick, Frederick, MD 21702). For
     information on which gene each clone represents, please
     visit our anonymous ftp site at
     ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
     a Note: this is a NIH_MGC Library."

ORIGIN
Query Match      68.7%; Score 20.6; DB 14; Length 804;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCGAGGTAGCACCAGT 27
    |||||
DB 99 TGGCGCTGAACCGAGGTAGCAGAGT 73
    |||||

RESULT 12
BZ773303
LOCUS
DEFINITION  mcv67d11.g1 HFOSMID007 Homo sapiens genomic, genomic survey
sequence.
ACCESSION  BZ773303
VERSION    BZ773303.1 GI:28946987
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```



```

/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2Pig"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

ORIGIN
Query Match      68.0%; Score 20.4; DB 10; Length 227;
Best Local Similarity 80.0%; Pred. No. 3.7e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  TGGCGGAGAACACGAGGAGTAGCACCAGTGG 30
Db      117 TGGAGCTGATACAGGAGAAAGCTCCAATGAG 88

RESULT 14
BQ136744/c
LOCUS      230 bp      mRNA      linear      EST 23-APR-2002
DEFINITION      NFO02D06ST1F1000 Developing stem Medicago truncatula cDNA clone
ACCESSION      BQ136744
VERSION      BQ136744.1 GI:20272868
KEYWORDS      EST.
SOURCE      Medicago truncatula (barrel medic)
ORGANISM      Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE      1 (bases 1 to 230)
AUTHORS      He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A.,
Bell,C.J., Flores,H.R., Inman,J.T., Weiler,J.W., May,G.D. and
Dixon,R.A.
TITLE      Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL      Medicago truncatula stem library
COMMENT      Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 230 Std Error: 0.00
Plate: 002 row: D column: 06
Seq primer: TCACACGAGGAACAGCTATGAC.
FEATURES
source      Location/Qualifiers
1..230
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone_lib="NFO02D06ST"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/clone_lib="Developing stem"
/notes="Vector: Lambda Zap; Contains a mixture of
internodal stem segments"

ORIGIN
Query Match      68.0%; Score 20.4; DB 13; Length 230;
Best Local Similarity 80.0%; Pred. No. 3.7e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  TGGCGGAGAACACGAGGAGTAGCACCAGTGG 30
Db      188 TGGTGAGAGCAGCAGCAGCAACCAATAG 159

RESULT 15

```


[illegible]

/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 293;
Best Local Similarity 80.0%; Pred. No. 4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGAACGAGGATGACCAATGAG 30
|||||
DB 179 TGGTCGAGAGCAGCAGCAACAATAAG 150

RESULT 20

BI267471/c

LOCUS BI267471 312 bp mRNA linear EST 18-JUL-2001
DEFINITION NF105E03IN1F1023 Insect herbivory Medicago truncatula cDNA clone
VERSION BI267471
KEYWORDS EST.

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 312)
Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula insect herbivory library
Unpublished (2000)
Contact: Korth K
Dept. of Plant Pathology
University of Arkansas
217 Plant Science Building, Fayetteville, AR 72701, USA
Tel: 501 575 5191
Fax: 501 575 7601
Email: korth@comp.uark.edu
Insert Length: 312 Std Error: 0.00
Plate: 105 row: E column: 03
Seq primer: TCACACGAGAACGCTATGAC.

FEATURES

source

1..312

/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF105E03IN"
/tissue_type="local and systemic leaves"
/dev_stages="mature"

/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 312;
Best Local Similarity 80.0%; Pred. No. 4.1e-03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGAACGAGGATGACCAATGAG 30
|||||
DB 188 TGGTCGAGAGCAGCAGCAACAATAAG 159

RESULT 21

BI265945/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 319)

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Korth K

Dept. of Plant Pathology

University of Arkansas

217 Plant Science Building, Fayetteville, AR 72701, USA

Tel: 501 575 5191

Fax: 501 575 7601

Email: korth@comp.uark.edu

Insert Length: 319 Std Error: 0.00

Plate: 092 row: F column: 12

Seq primer: TCACACGAGAACGCTATGAC.

Location/Qualifiers

1..319

source

/organism="Medicago truncatula"

/mol_type="mRNA"

/db_xref="taxon:3880"

/clone="NF092F12IN"

/tissue_type="local and systemic leaves"

/dev_stages="mature"

/clone_lib="Insect herbivory"

/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 319;

Best Local Similarity 80.0%; Pred. No. 4.1e+03;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGAACGAGGATGACCAATGAG 30

|||||

DB 197 TGGTCGAGAGCAGCAGCAACAATAAG 168

RESULT 22

AW127696/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 349)

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)

On Oct 25, 1999 this sequence version replaced gi:6115600.

Contact: Carroll P. Vance

BI265945 319 bp mRNA linear EST 18-JUL-2001
NF092F12IN1F1103 Insect herbivory Medicago truncatula cDNA clone
NF092F12IN 5', mRNA sequence.

ACCESSION BI265945

VERSION BI265945.1 GI:14869634

KEYWORDS EST.

SOURCE Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 319)

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Korth K

Dept. of Plant Pathology

University of Arkansas

217 Plant Science Building, Fayetteville, AR 72701, USA

Tel: 501 575 5191

Fax: 501 575 7601

Email: korth@comp.uark.edu

Insert Length: 319 Std Error: 0.00

Plate: 092 row: F column: 12

Seq primer: TCACACGAGAACGCTATGAC.

Location/Qualifiers

1..319

source

/organism="Medicago truncatula"

/mol_type="mRNA"

/db_xref="taxon:3880"

/clone="NF092F12IN"

/tissue_type="local and systemic leaves"

/dev_stages="mature"

/clone_lib="Insect herbivory"

/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: varcoe004@marcon.tc.umn.edu
Date: 9/15/99: Updated to the Database of Expressed Sequence Tags
(dbEST) on 04/27/00: More information is available at
'http://chrystle.tamu.edu/medicago'.
Seq primer: T3.

FEATURES

source
1. .349
Location/Qualifiers
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Genotype A17"
/db_xref="taxon:3880"
/clone="L270"
/tissue_type="leaves and cotyledons"
/dev_stage="mixture of cotyledons from five days old
plants and leaves obtained from two weeks old plants"
/lab_host="E. coli strain SOLR"
/clone_lib="DBLCL"
/note="Vector: pluescript SK +/-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA from the
mixture of cotyledons of five days old plants and leaves
of two weeks old plants. The cDNA was directionally
ligated into the Uni-ZAP XR vector from Stratagene and
packaged using Gigapack III Gold packaging extracts.
Plasmids containing cDNA inserts were excised from the
recombinant lambda-ZAP phage using EX-Assist helper phage
and propagated in SOLR cells."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 349;
Best Local Similarity 80.0%; Pred. No. 4.2e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TGGCGCAGACACGAGTAGCACCACATGAG 30
|||||
Db 215 TGGTGCAGACGAGCAGCAGCAACAATAAG 186

RESULT 23
BE323556/c
LOCUS
DEFINITION
BE323556
VERSION
KEYWORDS
SOURCE
ORGANISM
Medicago truncatula (barrel medic)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 358)
Lit, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula phosphate-starved leaf library
Unpublished (2000)
On Jul 14, 2000 this sequence version replaced gi:9197333.
Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Medicago Genome Initiative accession: MGI:S:19768
Insert Length: 794 Std Error: 0.00
Plate: 010 row: H column: 11

Seq primer: TCACACAGAAACAGCTATGAC.

FEATURES

source
1. .358
Location/Qualifiers
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF010H11PL"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/clone_lib="Phosphate starved leaf"
/note="Vector: Lambda Zap; At the trifoliolate stage, M.
truncatula plants were transplanted to phosphate-free sand
and grown for a further 30 days. During this 30 day
period, the plants were fertilized twice weekly with 1/2
Hoaglands solution containing only 20uM potassium
phosphate. RNA was prepared from above ground tissues."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 358;
Best Local Similarity 80.0%; Pred. No. 4.3e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TGGCGCAGACACGAGTAGCACCACATGAG 30
|||||
Db 203 TGGTGCAGACGAGCAGCAGCAACAATAAG 174

RESULT 24
BE315922/c
LOCUS
DEFINITION
BE315922
VERSION
KEYWORDS
SOURCE
ORGANISM
Medicago truncatula (barrel medic)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 361)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula leaf library
Unpublished (2000)
On Jul 14, 2000 this sequence version replaced gi:9189699.
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Medicago Genome Initiative accession: MGI:S:16476
Insert Length: 772 Std Error: 0.00
Plate: 028 row: C column: 04
Seq primer: TCACACAGAAACAGCTATGAC.

FEATURES

source
1. .361
Location/Qualifiers
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF028C04LF"
/tissue_type="leaf"
/dev_stage="Pooled developmental"
/clone_lib="Developing leaf"
/note="Vector: Lambda Zap; Contains a mixture of very
young, developing, mature and senescing leaves."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 361;

Email: kkorthe@comp.uark.edu
Medicago Genome Initiative accession: MGI:S:26212
Insert Length: 707 Std Error: 0.00
Plate: 025 row: C column: 04
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers

FEATURES
source

1..386
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF025C04IN"
/tissue_type="local and systemic leaves"
/dev_stage="mature"
/clone_lib="Insect herbivory"
/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 386;
Best Local Similarity 80.0%; Pred. No. 4.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30
|||
Db 286 TGGTGCAGAGCAGCAGCAGCAGCAACAATAAG 257

RESULT 28
BE323559/c

LOCUS
DEFINITION
BE323559
clone NF003F04PL1031 Phosphate starved leaf Medicago truncatula cDNA

ACCESSION
VERSION
BE323559.2 GI:11966494

KEYWORDS
SOURCE
ORGANISM

Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE

1 (bases 1 to 386)
Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula phosphate-starved leaf library
Unpublished (2000)
On Jul 14, 2000 this sequence version replaced gi:9197336.

JOURNAL
COMMENT

Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mj.harrison@noble.org
Medicago Genome Initiative accession: MGI:S:19994
Insert Length: 712 Std Error: 0.00
Plate: 003 row: F column: 04
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers

FEATURES
source

1..386
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF003F04PL"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/clone_lib="Phosphate starved leaf"
/note="Vector: Lambda Zap; At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand

and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 20mM potassium phosphate. RNA was prepared from above ground tissues."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 386;
Best Local Similarity 80.0%; Pred. No. 4.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30
|||
Db 181 TGGTGCAGAGCAGCAGCAGCAGCAACAATAAG 152

RESULT 29
BE249756/c

LOCUS
DEFINITION
BE249756
clone NF021H07LF1063 Developing leaf Medicago truncatula cDNA clone

ACCESSION
VERSION
BE249756.2 GI:11959451

KEYWORDS
SOURCE
ORGANISM

Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE

1 (bases 1 to 389)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula leaf library
Unpublished (2000)
On Jul 13, 2000 this sequence version replaced gi:9119823.

JOURNAL
COMMENT

Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Medicago Genome Initiative accession: MGI:S:15504
Insert Length: 722 Std Error: 0.00
Plate: 021 row: H column: 07
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers

FEATURES
source

1..389
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF021H07LF"
/tissue_type="leaf"
/dev_stage="Pooled developmental"
/clone_lib="Developing leaf"
/note="Vector: Lambda Zap; Contains a mixture of very young, developing, mature and senescing leaves."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 389;
Best Local Similarity 80.0%; Pred. No. 4.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30
|||
Db 158 TGGTGCAGAGCAGCAGCAGCAGCAACAATAAG 129

RESULT 30
BI263513/c

LOCUS
DEFINITION
BI263513
clone NF087A01PL1F1003 Phosphate starved leaf Medicago truncatula cDNA

clone NF087A01PL 5', mRNA sequence.
ACCESSION BI263513
VERSION BI263513.1 GI:14864818
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 398)
AUTHORS Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula phosphate-starved leaf library
COMMENT Unpublished (2000)
Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Insert Length: 398 Std Error: 0.00
Plate: 087 row: A column: 01
Seq primer: TCACACAGGAAACAGCTATGAC.

FEATURES
source
1..398
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF087A01PL"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/clone_lib="Phosphate starved leaf"
/note="Vector: Lambda Zap; At the trifoliolate stage, M.
truncatula plants were transplanted to phosphate-free sand
and grown for a further 30 days. During this 30 day
period, the plants were fertilized twice weekly with 1/2
Hoaglands solution containing only 20uM potassium
phosphate. RNA was prepared from above ground tissues."

ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 398;
Best Local Similarity 80.0%; Pred. No. 4.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCGAGTAGCACCACCAATGAG 30
|||
Db 228 TGGTCAGAGCAGCAGCAGCAACCAATAG 199
|||

Search completed: June 20, 2004, 14:15:59
Job time : 1425.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 33.1746 Seconds
(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgacagaccaggtagcaccacatgag 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*

4: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| C 1 | 30 | 100.0 | 4403765 | 3 | US-09-103-840A-2 |
| C 2 | 30 | 100.0 | 4411529 | 3 | US-09-103-840A-1 |
| C 3 | 20.6 | 68.7 | 510 | 4 | US-09-621-976-1444 |
| C 4 | 20.6 | 68.7 | 1476 | 4 | US-08-277-031B-10 |
| C 5 | 15.2 | 64.0 | 21706 | 4 | US-08-961-527-36 |
| C 6 | 18.8 | 62.7 | 1824 | 1 | US-08-305-505-1 |
| C 7 | 18.8 | 62.7 | 2907 | 4 | US-09-023-655-1053 |
| C 8 | 18.6 | 62.0 | 290 | 4 | US-09-313-294A-6733 |
| C 9 | 18.4 | 61.3 | 409 | 4 | US-09-621-976-18228 |
| C 10 | 18.4 | 61.3 | 461 | 4 | US-09-621-976-18229 |
| C 11 | 18.4 | 61.3 | 492 | 4 | US-08-621-976-3472 |
| C 12 | 18.4 | 61.3 | 538 | 4 | US-09-621-976-18226 |
| C 13 | 18.4 | 61.3 | 641 | 4 | US-09-621-976-18227 |
| C 14 | 18.4 | 61.3 | 4219 | 4 | US-09-976-594-752 |
| C 15 | 18.2 | 60.7 | 4403765 | 3 | US-09-103-840A-2 |
| C 16 | 18.2 | 60.7 | 4411529 | 3 | US-09-103-840A-1 |
| C 17 | 18 | 60.0 | 3879 | 3 | US-08-916-352-1 |
| C 18 | 18 | 60.0 | 16998 | 4 | US-08-676-610B-24 |
| C 19 | 18 | 60.0 | 197496 | 4 | US-09-877-177A-10 |
| C 20 | 18 | 60.0 | 1830121 | 4 | US-09-557-884-1 |
| C 21 | 18 | 60.0 | 1830121 | 4 | US-09-643-990A-1 |
| C 22 | 17.8 | 59.3 | 312 | 4 | US-08-973-544-2 |
| C 23 | 17.8 | 59.3 | 5581 | 4 | US-08-973-544-1 |
| C 24 | 17.8 | 59.3 | 5911 | 4 | US-08-311-174-4 |
| C 25 | 17.6 | 58.7 | 481 | 4 | US-08-621-976-1469 |
| C 26 | 17.6 | 58.7 | 654 | 4 | US-09-184-418C-99 |
| C 27 | 17.6 | 58.7 | 1527 | 4 | US-09-543-681A-2651 |

| | | | | | |
|------|------|--------|---|----------------------|--------------------|
| 17.6 | 58.7 | 1813 | 4 | US-09-620-312B-29 | Sequence 29, Appl |
| 17.6 | 58.7 | 9060 | 4 | US-09-184-418C-10 | Sequence 10, Appl |
| 17.4 | 58.0 | 948 | 4 | US-09-598-401C-107 | Sequence 107, Appl |
| 17.4 | 58.0 | 1766 | 4 | US-09-598-401C-117 | Sequence 117, Appl |
| 17.4 | 58.0 | 2283 | 4 | US-09-252-991A-3298 | Sequence 3298, Ap |
| 17.2 | 57.3 | 490 | 4 | US-09-621-976-18339 | Sequence 18339, A |
| 17.2 | 57.3 | 1266 | 4 | US-09-543-681A-285 | Sequence 285, App |
| 17.2 | 57.3 | 1902 | 4 | US-09-620-312B-1057 | Sequence 1057, Ap |
| 17.2 | 57.3 | 2571 | 4 | US-09-620-312B-302 | Sequence 302, App |
| 17.2 | 57.3 | 2598 | 4 | US-09-612-204B-23 | Sequence 23, Appl |
| 17.2 | 57.3 | 2830 | 2 | US-09-010-928B-1 | Sequence 1, Appli |
| 17.2 | 57.3 | 7015 | 4 | US-09-770-315-1 | Sequence 1, Appli |
| 17.2 | 57.3 | 80161 | 3 | US-09-036-987A-1 | Sequence 1, Appli |
| 17.2 | 57.3 | 80161 | 3 | US-09-370-700-1 | Sequence 1, Appli |
| 17.2 | 57.3 | 80161 | 4 | US-09-603-207-1 | Sequence 1, Appli |
| 17 | 56.7 | 4404 | 4 | US-09-621-976-10068 | Sequence 10068, A |
| 17 | 56.7 | 445 | 4 | US-09-621-976-18828 | Sequence 18828, A |
| 17 | 56.7 | 717 | 4 | US-09-134-001C-299 | Sequence 299, App |
| 17 | 56.7 | 1098 | 4 | US-09-328-352-1981 | Sequence 1981, Ap |
| 17 | 56.7 | 1509 | 4 | US-09-134-001C-761 | Sequence 761, App |
| 17 | 56.7 | 2634 | 2 | US-08-507-166-7 | Sequence 7, Appli |
| 17 | 56.7 | 2634 | 4 | US-09-391-340-7 | Sequence 7, Appli |
| 17 | 56.7 | 2866 | 4 | US-09-724-864-30 | Sequence 30, Appl |
| 17 | 56.7 | 3651 | 2 | US-08-790-374-1 | Sequence 1, Appli |
| 17 | 56.7 | 9775 | 3 | US-08-977-171-1 | Sequence 1, Appli |
| 16.8 | 56.0 | 668 | 3 | US-08-998-416-150 | Sequence 150, App |
| 16.8 | 56.0 | 705 | 4 | US-09-540-236-278 | Sequence 278, App |
| 16.8 | 56.0 | 909 | 4 | US-08-148-708-9 | Sequence 9, Appli |
| 16.8 | 56.0 | 1179 | 4 | US-08-148-708-3 | Sequence 3, Appli |
| 16.8 | 56.0 | 1378 | 1 | US-08-759-848-2 | Sequence 2, Appli |
| 16.8 | 56.0 | 1378 | 4 | US-08-148-708-6 | Sequence 6, Appli |
| 16.8 | 56.0 | 1378 | 4 | US-09-016-434-1123 | Sequence 1123, Ap |
| 16.8 | 56.0 | 1378 | 5 | PCT-US95-09383-2 | Sequence 2, Appli |
| 16.8 | 56.0 | 2205 | 3 | US-08-687-590-59 | Sequence 59, Appl |
| 16.8 | 56.0 | 2574 | 2 | US-08-677-734A-8 | Sequence 8, Appli |
| 16.8 | 56.0 | 2574 | 4 | US-09-097-053-8 | Sequence 8, Appli |
| 16.8 | 56.0 | 3002 | 4 | US-09-369-364A-1 | Sequence 1, Appli |
| 16.8 | 56.0 | 3577 | 4 | US-09-620-312B-457 | Sequence 457, App |
| 16.8 | 56.0 | 6703 | 4 | US-09-596-002-7 | Sequence 7, Appli |
| 16.8 | 56.0 | 6703 | 4 | US-08-545-528B-1 | Sequence 1, Appli |
| 16.8 | 56.0 | 590073 | 3 | US-08-865-960-4 | Sequence 4, Appli |
| 16.6 | 55.3 | 39 | 3 | US-09-252-991A-15319 | Sequence 15319, A |
| 16.6 | 55.3 | 396 | 4 | US-08-346-611-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 402 | 1 | US-08-346-611-3 | Sequence 3, Appli |
| 16.6 | 55.3 | 402 | 1 | US-08-794-494-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 402 | 2 | US-08-794-494-3 | Sequence 3, Appli |
| 16.6 | 55.3 | 456 | 4 | US-09-252-991A-15319 | Sequence 15319, A |
| 16.6 | 55.3 | 891 | 4 | US-09-134-001C-1636 | Sequence 1636, Ap |
| 16.6 | 55.3 | 945 | 4 | US-09-252-991A-15367 | Sequence 15367, A |
| 16.6 | 55.3 | 1047 | 4 | US-09-784-508-15 | Sequence 15, Appl |
| 16.6 | 55.3 | 1074 | 4 | US-09-252-991A-4289 | Sequence 4289, Ap |
| 16.6 | 55.3 | 1305 | 4 | US-09-242-859A-3 | Sequence 3, Appli |
| 16.6 | 55.3 | 1305 | 4 | US-09-242-859A-7 | Sequence 7, Appli |
| 16.6 | 55.3 | 1411 | 4 | US-09-370-838-154 | Sequence 154, App |
| 16.6 | 55.3 | 1426 | 4 | US-09-252-991A-4561 | Sequence 4561, Ap |
| 16.6 | 55.3 | 1716 | 4 | US-09-252-991A-15268 | Sequence 15268, A |
| 16.6 | 55.3 | 1848 | 4 | US-09-252-991A-15408 | Sequence 15408, A |
| 16.6 | 55.3 | 1999 | 4 | US-09-716-129-44 | Sequence 44, Appl |
| 16.6 | 55.3 | 2028 | 4 | US-09-489-039A-2010 | Sequence 2010, Ap |
| 16.6 | 55.3 | 2159 | 3 | US-08-286-870A-7 | Sequence 7, Appli |
| 16.6 | 55.3 | 2368 | 2 | US-08-520-933-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 2368 | 4 | US-09-285-940-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 2368 | 6 | 5262177-1 | Patent No. 5262177 |
| 16.6 | 55.3 | 3293 | 2 | US-08-442-809A-75 | Sequence 75, Appl |
| 16.6 | 55.3 | 3909 | 4 | US-08-961-527-12 | Sequence 12, Appl |
| 16.6 | 55.3 | 14113 | 3 | US-09-223-134-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 14113 | 3 | US-08-992-801-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 14113 | 3 | US-09-223-535-1 | Sequence 1, Appli |
| 16.6 | 55.3 | 176373 | 3 | US-09-128-153-17 | Sequence 17, Appl |
| 16.4 | 54.7 | 27 | 4 | US-09-011-336-39 | Sequence 39, Appl |
| 16.4 | 54.7 | 60 | 3 | US-08-290-736C-37 | Sequence 37, Appl |
| 16.4 | 54.7 | 71 | 3 | US-08-290-736C-46 | Sequence 46, Appl |
| 16.4 | 54.7 | 71 | 3 | US-08-290-736C-47 | Sequence 47, Appl |

| | | | | | | | | | | | | | |
|-------|------|------|------|---|---------------------|--------------------|-------|------|------|-------|---|---------------------|--------------------|
| C 101 | 16.4 | 54.7 | 122 | 3 | US-09-106-182-12 | Sequence 12, Appl | C 174 | 16.4 | 54.7 | 2172 | 3 | US-09-267-439-26 | Sequence 26, Appl |
| C 102 | 16.4 | 54.7 | 129 | 4 | US-09-589-287B-16 | Sequence 16, Appl | C 175 | 16.4 | 54.7 | 2172 | 4 | US-09-073-138-26 | Sequence 26, Appl |
| C 103 | 16.4 | 54.7 | 129 | 4 | US-09-268-311-10 | Sequence 10, Appl | C 176 | 16.4 | 54.7 | 2343 | 4 | US-09-328-352-4054 | Sequence 4054, Ap |
| C 104 | 16.4 | 54.7 | 129 | 4 | US-09-588-947A-16 | Sequence 16, Appl | C 177 | 16.4 | 54.7 | 2425 | 4 | US-09-023-655-234 | Sequence 234, App |
| C 105 | 16.4 | 54.7 | 129 | 4 | US-09-154-219-10 | Sequence 10, Appl | C 178 | 16.4 | 54.7 | 2790 | 4 | US-09-254-594-5 | Sequence 5, Appli |
| C 106 | 16.4 | 54.7 | 129 | 4 | US-09-589-286A-16 | Sequence 16, Appl | C 179 | 16.4 | 54.7 | 3185 | 1 | US-08-117-362-2 | Sequence 2, Appli |
| C 107 | 16.4 | 54.7 | 129 | 4 | US-09-489-039A-2612 | Sequence 2612, Ap | C 180 | 16.4 | 54.7 | 3185 | 1 | US-08-486-924-2 | Sequence 2, Appli |
| C 108 | 16.4 | 54.7 | 217 | 4 | US-09-621-976-12915 | Sequence 12915, A | C 181 | 16.4 | 54.7 | 3185 | 1 | US-08-486-929A-2 | Sequence 2, Appli |
| C 109 | 16.4 | 54.7 | 297 | 4 | US-09-621-976-13157 | Sequence 13157, A | C 182 | 16.4 | 54.7 | 3432 | 4 | US-09-254-594-4 | Sequence 4, Appli |
| C 110 | 16.4 | 54.7 | 299 | 4 | US-09-621-976-12708 | Sequence 12708, A | C 183 | 16.4 | 54.7 | 3914 | 1 | US-08-117-373-11 | Sequence 11, Appl |
| C 111 | 16.4 | 54.7 | 314 | 4 | US-09-223-575-77 | Sequence 77, Appl | C 184 | 16.4 | 54.7 | 4104 | 1 | US-07-998-003A-94 | Sequence 94, Appl |
| C 112 | 16.4 | 54.7 | 314 | 4 | US-09-389-681-77 | Sequence 77, Appl | C 185 | 16.4 | 54.7 | 4104 | 1 | US-08-453-274B-94 | Sequence 94, Appl |
| C 113 | 16.4 | 54.7 | 314 | 4 | US-09-620-405B-77 | Sequence 77, Appl | C 186 | 16.4 | 54.7 | 4104 | 1 | US-08-453-695A-94 | Sequence 94, Appl |
| C 114 | 16.4 | 54.7 | 314 | 4 | US-09-339-338-77 | Sequence 77, Appl | C 187 | 16.4 | 54.7 | 4104 | 1 | US-08-268-161A-94 | Sequence 94, Appl |
| C 115 | 16.4 | 54.7 | 314 | 4 | US-09-433-826B-77 | Sequence 77, Appl | C 188 | 16.4 | 54.7 | 4104 | 2 | US-08-453-702A-94 | Sequence 94, Appl |
| C 116 | 16.4 | 54.7 | 314 | 4 | US-09-604-287A-77 | Sequence 77, Appl | C 189 | 16.4 | 54.7 | 4104 | 3 | US-09-099-639-94 | Sequence 94, Appl |
| C 117 | 16.4 | 54.7 | 314 | 4 | US-09-285-480-77 | Sequence 77, Appl | C 190 | 16.4 | 54.7 | 4104 | 5 | PCT-US93-12588-94 | Sequence 94, Appl |
| C 118 | 16.4 | 54.7 | 314 | 4 | US-09-834-759-77 | Sequence 77, Appl | C 191 | 16.4 | 54.7 | 4286 | 4 | PCT-US95-08071-94 | Sequence 94, Appl |
| C 119 | 16.4 | 54.7 | 321 | 4 | US-09-195-106-4 | Sequence 4, Appli | C 192 | 16.4 | 54.7 | 4286 | 4 | US-09-376-594-632 | Sequence 632, App |
| C 120 | 16.4 | 54.7 | 353 | 4 | US-09-643-597-138 | Sequence 138, App | C 193 | 16.4 | 54.7 | 4650 | 1 | US-07-998-003A-102 | Sequence 102, App |
| C 121 | 16.4 | 54.7 | 353 | 4 | US-09-480-884A-138 | Sequence 138, App | C 194 | 16.4 | 54.7 | 4650 | 1 | US-08-453-274B-102 | Sequence 102, App |
| C 122 | 16.4 | 54.7 | 353 | 4 | US-09-542-615A-138 | Sequence 138, App | C 195 | 16.4 | 54.7 | 4650 | 1 | US-08-453-695A-102 | Sequence 102, App |
| C 123 | 16.4 | 54.7 | 353 | 4 | US-09-606-421B-138 | Sequence 138, App | C 196 | 16.4 | 54.7 | 4650 | 2 | US-08-268-161A-102 | Sequence 102, App |
| C 124 | 16.4 | 54.7 | 353 | 4 | US-09-221-107-138 | Sequence 138, App | C 197 | 16.4 | 54.7 | 4650 | 2 | US-08-453-702A-102 | Sequence 102, App |
| C 125 | 16.4 | 54.7 | 366 | 4 | US-09-329-884-5 | Sequence 5, Appli | C 198 | 16.4 | 54.7 | 4650 | 5 | US-09-099-639-102 | Sequence 102, App |
| C 126 | 16.4 | 54.7 | 392 | 3 | US-09-030-607-194 | Sequence 194, App | C 199 | 16.4 | 54.7 | 4650 | 5 | PCT-US93-12588-102 | Sequence 102, App |
| C 127 | 16.4 | 54.7 | 392 | 4 | US-09-439-313-194 | Sequence 194, App | C 200 | 16.4 | 54.7 | 4650 | 5 | PCT-US95-08071-102 | Sequence 102, App |
| C 128 | 16.4 | 54.7 | 392 | 4 | US-09-352-616A-194 | Sequence 194, App | C 201 | 16.4 | 54.7 | 4680 | 1 | US-08-234-358-1 | Sequence 1, Appli |
| C 129 | 16.4 | 54.7 | 392 | 4 | US-09-232-149A-194 | Sequence 194, App | C 202 | 16.4 | 54.7 | 4680 | 1 | US-08-475-391-1 | Sequence 1, Appli |
| C 130 | 16.4 | 54.7 | 392 | 4 | US-09-159-813-194 | Sequence 194, App | C 203 | 16.4 | 54.7 | 4680 | 2 | US-08-709-609-1 | Sequence 1, Appli |
| C 131 | 16.4 | 54.7 | 392 | 4 | US-09-636-215-194 | Sequence 194, App | C 204 | 16.4 | 54.7 | 4680 | 5 | PCT-US95-07178-1 | Sequence 3, Appli |
| C 132 | 16.4 | 54.7 | 392 | 4 | US-09-685-166A-194 | Sequence 194, App | C 205 | 16.4 | 54.7 | 4771 | 3 | US-08-940-062-3 | Patent No. 549251 |
| C 133 | 16.4 | 54.7 | 392 | 4 | US-09-115-453-194 | Sequence 194, App | C 206 | 16.4 | 54.7 | 4825 | 6 | 5459251-1 | Sequence 2, Appli |
| C 134 | 16.4 | 54.7 | 392 | 4 | US-09-688-489-194 | Sequence 194, App | C 207 | 16.4 | 54.7 | 4910 | 2 | US-08-331-384-2 | Sequence 2, Appli |
| C 135 | 16.4 | 54.7 | 504 | 4 | US-09-329-884-21 | Sequence 21, Appl | C 208 | 16.4 | 54.7 | 4910 | 2 | US-08-836-087-2 | Sequence 2, Appli |
| C 136 | 16.4 | 54.7 | 504 | 4 | US-09-329-884-19 | Sequence 19, Appl | C 209 | 16.4 | 54.7 | 4910 | 3 | US-09-246-320-2 | Sequence 2, Appli |
| C 137 | 16.4 | 54.7 | 510 | 3 | US-08-432-994A-7 | Sequence 7, Appli | C 210 | 16.4 | 54.7 | 4910 | 4 | US-09-546-738-2 | Sequence 2, Appli |
| C 138 | 16.4 | 54.7 | 510 | 4 | US-08-250-846-7 | Sequence 7, Appli | C 211 | 16.4 | 54.7 | 5086 | 2 | US-08-465-485A-19 | Sequence 19, Appl |
| C 139 | 16.4 | 54.7 | 546 | 4 | US-09-134-000C-1443 | Sequence 1443, Ap | C 212 | 16.4 | 54.7 | 5086 | 3 | US-08-365-486A-14 | Sequence 14, Appl |
| C 140 | 16.4 | 54.7 | 640 | 4 | US-09-758-282B-17 | Sequence 17, Appl | C 213 | 16.4 | 54.7 | 5086 | 4 | US-09-080-285-19 | Sequence 19, Appl |
| C 141 | 16.4 | 54.7 | 650 | 4 | US-09-235-451-7 | Sequence 7, Appli | C 214 | 16.4 | 54.7 | 5086 | 4 | US-08-380-342-14 | Sequence 14, Appl |
| C 142 | 16.4 | 54.7 | 653 | 3 | US-08-998-416-149 | Sequence 149, App | C 215 | 16.4 | 54.7 | 5086 | 4 | US-09-124-426-19 | Sequence 19, Appl |
| C 143 | 16.4 | 54.7 | 654 | 4 | US-09-328-352-2942 | Sequence 2942, Ap | C 216 | 16.4 | 54.7 | 5086 | 4 | US-09-233-327-7 | Sequence 7, Appli |
| C 144 | 16.4 | 54.7 | 657 | 4 | US-09-023-655-1261 | Sequence 1261, Ap | C 217 | 16.4 | 54.7 | 5086 | 5 | PCT-US93-05651-4 | Sequence 4, Appli |
| C 145 | 16.4 | 54.7 | 734 | 4 | US-09-692-081-9 | Sequence 9, Appli | C 218 | 16.4 | 54.7 | 5086 | 5 | PCT-US93-06251-2 | Sequence 2, Appli |
| C 146 | 16.4 | 54.7 | 810 | 4 | US-09-328-352-2013 | Sequence 2013, Ap | C 219 | 16.4 | 54.7 | 5094 | 4 | US-09-234-186-7 | Patent No. 5506344 |
| C 147 | 16.4 | 54.7 | 840 | 4 | US-09-328-352-591 | Sequence 591, App | C 220 | 16.4 | 54.7 | 5104 | 6 | 5506344-1 | Sequence 1, Appli |
| C 148 | 16.4 | 54.7 | 956 | 4 | US-09-220-132-162 | Sequence 162, App | C 221 | 16.4 | 54.7 | 7214 | 4 | US-09-438-268-1 | Sequence 1, Appli |
| C 149 | 16.4 | 54.7 | 1053 | 3 | US-09-107-532A-16 | Sequence 16, Appl | C 222 | 16.4 | 54.7 | 7379 | 4 | US-09-341-587-5 | Sequence 5, Appli |
| C 150 | 16.4 | 54.7 | 1128 | 4 | US-08-795-473B-1 | Sequence 1, Appli | C 223 | 16.4 | 54.7 | 7557 | 4 | US-09-770-315-3 | Sequence 3, Appli |
| C 151 | 16.4 | 54.7 | 1128 | 4 | US-09-439-856-1 | Sequence 1, Appli | C 224 | 16.4 | 54.7 | 8067 | 1 | US-08-117-362-8 | Sequence 8, Appli |
| C 152 | 16.4 | 54.7 | 1128 | 4 | US-09-866-344-30 | Sequence 30, Appl | C 225 | 16.4 | 54.7 | 8067 | 1 | US-08-486-924-8 | Sequence 8, Appli |
| C 153 | 16.4 | 54.7 | 1128 | 4 | US-09-626-896-15 | Sequence 15, Appl | C 226 | 16.4 | 54.7 | 8151 | 4 | US-08-486-929A-8 | Sequence 2, Appli |
| C 154 | 16.4 | 54.7 | 1145 | 6 | 5510472-1 | Patent No. 5510472 | C 227 | 16.4 | 54.7 | 8179 | 4 | US-09-438-268-5 | Sequence 5, Appli |
| C 155 | 16.4 | 54.7 | 1155 | 4 | US-09-970-516-1 | Sequence 1, Appli | C 228 | 16.4 | 54.7 | 8179 | 4 | US-09-770-315-2 | Sequence 2, Appli |
| C 156 | 16.4 | 54.7 | 1192 | 4 | US-09-692-081-1 | Sequence 1, Appli | C 229 | 16.4 | 54.7 | 8698 | 4 | US-09-073-492-1 | Sequence 1, Appli |
| C 157 | 16.4 | 54.7 | 1215 | 4 | US-09-489-039A-1714 | Sequence 1714, Ap | C 230 | 16.4 | 54.7 | 15462 | 3 | US-08-976-063E-1 | Sequence 9, Appli |
| C 158 | 16.4 | 54.7 | 1404 | 4 | US-08-328-352-3384 | Sequence 3384, Ap | C 231 | 16.4 | 54.7 | 32679 | 4 | US-09-590-759-10 | Sequence 10, Appl |
| C 159 | 16.4 | 54.7 | 1446 | 4 | US-08-976-063B-27 | Sequence 27, Appl | C 232 | 16.2 | 54.0 | 154 | 4 | US-09-313-294A-711 | Sequence 711, App |
| C 160 | 16.4 | 54.7 | 1533 | 4 | US-09-205-258-90 | Sequence 90, Appl | C 233 | 16.2 | 54.0 | 154 | 4 | US-09-313-294A-711 | Sequence 711, App |
| C 161 | 16.4 | 54.7 | 1609 | 4 | US-09-376-594-850 | Sequence 850, App | C 234 | 16.2 | 54.0 | 262 | 4 | US-09-634-137-23 | Sequence 23, Appl |
| C 162 | 16.4 | 54.7 | 1708 | 4 | US-09-919-172-30 | Sequence 30, Appl | C 235 | 16.2 | 54.0 | 287 | 4 | US-08-990-823-111 | Sequence 111, App |
| C 163 | 16.4 | 54.7 | 1709 | 2 | US-09-010-398-2 | Sequence 2, Appli | C 236 | 16.2 | 54.0 | 431 | 3 | US-09-477-135A-111 | Sequence 111, App |
| C 164 | 16.4 | 54.7 | 1709 | 3 | US-09-366-260-2 | Sequence 11, Appl | C 237 | 16.2 | 54.0 | 431 | 3 | US-09-477-135A-111 | Sequence 111, App |
| C 165 | 16.4 | 54.7 | 1709 | 3 | US-09-153-804-11 | Sequence 11, Appl | C 238 | 16.2 | 54.0 | 457 | 4 | US-09-590-759-9 | Sequence 9, Appli |
| C 166 | 16.4 | 54.7 | 1784 | 4 | US-09-023-655-891 | Sequence 891, App | C 239 | 16.2 | 54.0 | 457 | 4 | US-09-590-759-9 | Sequence 9, Appli |
| C 167 | 16.4 | 54.7 | 1874 | 4 | US-08-388-852B-1 | Sequence 1, Appli | C 240 | 16.2 | 54.0 | 492 | 4 | US-09-252-991A-8970 | Sequence 8970, Ap |
| C 168 | 16.4 | 54.7 | 2115 | 3 | US-09-056-105-1 | Sequence 1, Appli | C 241 | 16.2 | 54.0 | 492 | 4 | US-09-085-199B-36 | Sequence 36, Appl |
| C 169 | 16.4 | 54.7 | 2130 | 3 | US-08-434-784B-91 | Sequence 91, Appl | C 242 | 16.2 | 54.0 | 522 | 4 | US-09-252-991A-8970 | Sequence 8970, Ap |
| C 170 | 16.4 | 54.7 | 2131 | 2 | US-08-417-174-26 | Sequence 26, Appl | C 243 | 16.2 | 54.0 | 522 | 4 | US-09-252-991A-8970 | Sequence 8970, Ap |
| C 171 | 16.4 | 54.7 | 2172 | 2 | US-08-231-565A-26 | Sequence 26, Appl | C 244 | 16.2 | 54.0 | 544 | 4 | US-09-621-976-2695 | Sequence 2695, Ap |
| C 172 | 16.4 | 54.7 | 2172 | 2 | US-08-007-361-26 | Sequence 26, Appl | C 245 | 16.2 | 54.0 | 636 | 4 | US-09-252-991A-8970 | Sequence 8970, Ap |
| C 173 | 16.4 | 54.7 | 2172 | 2 | | | C 246 | 16.2 | 54.0 | 771 | 2 | US-08-208-005C-1 | Sequence 1, Appli |

247 16.2 54.0 771 2 US-09-038-597A-1 Sequence 1, Appli
248 16.2 54.0 771 2 US-08-431-117A-1 Sequence 1, Appli
249 16.2 54.0 780 3 US-08-855-825-11 Sequence 11, Appl
250 16.2 54.0 780 3 US-08-855-825-13 Sequence 13, Appl
251 16.2 54.0 780 3 US-09-252-991A-9118 Sequence 9118, Ap
252 16.2 54.0 792 1 US-08-578-709-10 Sequence 10, Appl
253 16.2 54.0 984 4 US-09-221-017B-708 Sequence 708, App
254 16.2 54.0 1003 4 US-09-634-137-31 Sequence 31, Appl
255 16.2 54.0 1089 4 US-09-107-532A-3439 Sequence 3439, Ap
256 16.2 54.0 1122 4 US-09-252-991A-12729 Sequence 12729, A
257 16.2 54.0 1152 4 US-09-252-991A-12886 Sequence 12886, A
258 16.2 54.0 1182 4 US-09-107-532A-58 Sequence 58, Appl
259 16.2 54.0 1316 4 US-09-461-325-91 Sequence 91, Appl
260 16.2 54.0 1325 4 US-10-012-542-91 Sequence 91, Appl
261 16.2 54.0 1365 4 US-09-023-655-1343 Sequence 1343, Ap
262 16.2 54.0 1365 4 US-09-328-352-9222 Sequence 922, App
263 16.2 54.0 1431 4 US-09-252-991A-10507 Sequence 10507, A
264 16.2 54.0 1547 4 US-09-620-312D-966 Sequence 966, App
265 16.2 54.0 1566 4 US-09-252-991A-8859 Sequence 8859, Ap
266 16.2 54.0 1619 4 US-09-522-714-11 Sequence 11, Appl
267 16.2 54.0 1674 4 US-09-489-039A-3633 Sequence 3633, Ap
268 16.2 54.0 1866 4 US-09-252-991A-3439 Sequence 3439, Ap
269 16.2 54.0 1897 4 US-09-556-870A-1 Sequence 1, Appli
270 16.2 54.0 1845 4 US-09-540-236-938 Sequence 938, App
271 16.2 54.0 1878 4 US-09-134-000C-667 Sequence 667, App
272 16.2 54.0 1977 4 US-08-578-709-14 Sequence 14, Appl
273 16.2 54.0 2028 4 US-09-252-991A-12500 Sequence 12500, A
274 16.2 54.0 2040 4 US-09-252-991A-12977 Sequence 12977, A
275 16.2 54.0 2079 4 US-09-489-039A-3593 Sequence 3593, Ap
276 16.2 54.0 2088 4 US-09-458-481B-3 Sequence 3, Appli
277 16.2 54.0 2190 4 US-09-625-188-19 Sequence 19, Appl
278 16.2 54.0 2301 1 US-08-306-691B-23 Sequence 23, Appl
279 16.2 54.0 2301 4 US-09-167-206-3 Sequence 3, Appli
280 16.2 54.0 2301 5 PCT-US93-06251-78 Sequence 78, Appl
281 16.2 54.0 2697 4 US-09-620-312D-401 Sequence 401, App
282 16.2 54.0 2858 4 US-09-688-078-5 Sequence 5, Appli
283 16.2 54.0 3098 3 US-09-232-200-58 Sequence 58, Appl
284 16.2 54.0 3098 4 US-09-232-197-58 Sequence 58, Appl
285 16.2 54.0 3098 4 US-09-232-201-58 Sequence 58, Appl
286 16.2 54.0 3098 4 US-09-443-184-43 Sequence 43, Appl
287 16.2 54.0 3136 4 US-09-232-195-58 Sequence 58, Appl
288 16.2 54.0 3757 4 US-09-023-655-1502 Sequence 1502, Ap
289 16.2 54.0 3999 4 US-09-252-991A-1201 Sequence 9201, Ap
290 16.2 54.0 4095 4 US-09-252-991A-10309 Sequence 10309, A
291 16.2 54.0 4171 1 US-08-308-881-5 Sequence 5, Appli
292 16.2 54.0 4171 2 US-09-058-263-5 Sequence 5, Appli
293 16.2 54.0 4171 2 US-09-059-099-5 Sequence 5, Appli
294 16.2 54.0 4171 3 US-09-058-264-5 Sequence 5, Appli
295 16.2 54.0 4171 4 US-09-455-962-5 Sequence 5, Appli
296 16.2 54.0 4171 5 PCT-US93-06530-5 Sequence 5, Appli
297 16.2 54.0 6426 4 US-09-976-594-136 Sequence 136, App
298 16.2 54.0 9633 6 5223423-1 Patent No. 5223423
299 16.2 54.0 148567 4 US-09-801-876B-3 Sequence 3, Appli
300 16.2 54.0 148567 4 US-10-254-869-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

Query Match 100.0%; Score 30; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30
Db 24986 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 24957

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

Query Match 100.0%; Score 30; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30
Db 25004 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 24975

RESULT 3
US-09-621-976-1444/c
; Sequence 1444, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1444
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS

```

; LOCATION: 58..363
; NAME/KEY: sig_peptide
; LOCATION: 58..111
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 7.6999980926514
; OTHER INFORMATION: seq LLFLALLTGLLLL/LV
US-09-621-976-1444

Query Match
Best Local Similarity 58.7%; Score 20.6; DB 4; Length 510;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 125 TGGCGCTGAACCCAGGAGTAGCAGAGCT 99

RESULT 4
US-08-277-031B-10/c
; Sequence 10, Application US/08277031B
; Patent No. 6620593
; GENERAL INFORMATION:
; APPLICANT: Hayashi, Koji
; APPLICANT: Sakaki, Toshiyuki
; APPLICANT: Yabusaki, Yoshiyasu
; APPLICANT: Komai, Koichiro
; APPLICANT: Kaneko, Hideo
; APPLICANT: Nakatsuka, Iwao
; TITLE OF INVENTION: METHOD FOR SAFETY EVALUATION OF
; TITLE OF INVENTION: CHEMICAL COMPOUND USING RECOMBINANT YEAST EXPRESSING
; TITLE OF INVENTION: HUMAN CYTOCHROME P450
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5-inch, 1.44MB
; COMPUTER: IBM PC
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,031B
; FILING DATE: 19-JULY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-201120/1993
; APPLICATION NUMBER: JP-180246/1993
; APPLICATION NUMBER: JP-208279/1993
; FILING DATE: 20-07-1993
; FILING DATE: 21-07-1993
; FILING DATE: 30-07-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Raymond C. Stewart
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 20-3530P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1476
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-277-031B-10

Query Match
Best Local Similarity 68.7%; Score 20.6; DB 4; Length 1476;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 125 TGGCGCTGAACCCAGGAGTAGCAGAGCT 99

RESULT 5
US-08-961-527-36
; Sequence 36, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-36

Query Match
Best Local Similarity 64.0%; Score 19.2; DB 4; Length 21706;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCACCACATG 28
Db 6961 GAAGAACCCAGGAGTAGCTCCATG 6984

RESULT 6
US-08-305-505-1/c
; Sequence 1, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Mizutoko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
```

```

; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patenclin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-305-505-1

Query Match 62.7%; Score 18.8; DB 1; Length 1824;
Best Local Similarity 90.9%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCGCAGACGAGGAGTAGCACC 24
Db 1196 GCGCAGACGAGGAGTAGCACC 1175

RESULT 7
US-09-023-655-1053/c
; Sequence 1053, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071

; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patenclin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-305-505-1

Query Match 62.7%; Score 18.8; DB 1; Length 1824;
Best Local Similarity 90.9%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCGCAGACGAGGAGTAGCACC 24
Db 1196 GCGCAGACGAGGAGTAGCACC 1175

RESULT 7
US-09-023-655-1053/c
; Sequence 1053, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071

; COUNTRY: U.S.A.
; ZIP: 53202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1053:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2907 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g181293
; US-09-023-655-1053

Query Match 62.7%; Score 18.8; DB 4; Length 2907;
Best Local Similarity 90.9%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACGAGGAGTAGCA 22
Db 74 TGACGCTGACGAGGAGTAGCA 53

RESULT 8
US-09-313-294A-6733/c
; Sequence 6733, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6733
; LENGTH: 290
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700352309H1
; NAME/KEY: unsure
; LOCATION: 3, 14, 103-104
; OTHER INFORMATION: a, t, c, g, or other
; US-09-313-294A-6733

Query Match 62.0%; Score 18.6; DB 4; Length 290;
Best Local Similarity 84.0%; Pred. No. 54;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCAGAACGAGGAGTAGCACCAGTGA 29
Db 132 GTAGAACGAGTAGAAGCACCAGTCA 108

RESULT 9
US-09-621-976-18228
; Sequence 18228, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.EW
```

```
; SEQ ID NO 18228
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 55
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18228

Query Match          61.3%; Score 18.4; DB 4; Length 409;
Best Local Similarity 78.6%; Pred. No. 71;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 237 TGGCCAGCAGCAGGACTGGCGGCAATG 264

RESULT 10
US-09-621-976-18229
; Sequence 18229, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18229
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18229

Query Match          61.3%; Score 18.4; DB 4; Length 461;
Best Local Similarity 78.6%; Pred. No. 72;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 238 TGGCCAGCAGCAGGACTGGCGGCAATG 265

RESULT 11
US-09-621-976-3472
; Sequence 3472, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3472
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 156..347
US-09-621-976-3472

Query Match          61.3%; Score 18.4; DB 4; Length 492;
Best Local Similarity 78.6%; Pred. No. 73;
```

```
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 289 TGGCCAGCAGCAGGACTGGCGGCAATG 316

RESULT 12
US-09-621-976-18226
; Sequence 18226, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18226
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18226

Query Match          61.3%; Score 18.4; DB 4; Length 538;
Best Local Similarity 78.6%; Pred. No. 75;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 336 TGGCCAGCAGCAGGACTGGCGGCAATG 363

RESULT 13
US-09-621-976-18227
; Sequence 18227, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18227
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18227

Query Match          61.3%; Score 18.4; DB 4; Length 641;
Best Local Similarity 78.6%; Pred. No. 77;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 297 TGGCCAGCAGCAGGACTGGCGGCAATG 324

RESULT 14
US-09-976-594-752
; Sequence 752, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
```

```
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 752
; LENGTH: 4219
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6573549 1330149.46
; NAME/KEY: unsure
; LOCATION: 1565-1703, 3787, 3883
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-752

Query Match 61.3%; Score 18.4; DB 4; Length 4219;
Best Local Similarity 78.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GCGCAGACAGGAGTAGCACCATGA 29
DB 134 GGAGCAGATCCAGGAGAGCTCATATGA 161

RESULT 15
US-09-103-840A-2
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; TYPE: DNA
; LENGTH: 4403765
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 60.7%; Score 18.2; DB 3; Length 4403765;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGCAGACAGGAGTAGCACC 24
DB 3660571 GCGCAGACAGGAGTAGGTACC 3660593

RESULT 16
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
```

```
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 60.7%; Score 18.2; DB 3; Length 4411529;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGCAGACAGGAGTAGCACC 24
DB 3665785 GCGCAGACAGGAGTAGGTACC 3665807

RESULT 17
US-08-916-352-1
; Sequence 1, Application US/08916352
; Patent No. 6166191
; GENERAL INFORMATION:
; APPLICANT: CHIRON CORPORATION
; TITLE OF INVENTION: HUMAN POLYHOMEOTIC 1 (hph1) ACTS AS A
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/916,352
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: POTTER, JANE
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1355.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 1
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-916-352-1

Query Match 60.8%; Score 18; DB 3; Length 3879;
Best Local Similarity 80.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACC 26
DB 1268 TGGCGCAGCAGCAGCAGCAGCA 1293

RESULT 18
```

US-09-676-610B-24/c
; Sequence 24, Application US/09676610B
; Patent No. 644465
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Susan M. Preier
; TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
; FILE REFERENCE: RTS-0138
; CURRENT APPLICATION NUMBER: US/09/676,610B
; CURRENT FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 182
; SEQ ID NO 24
; LENGTH: 169998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1208)...(1472)
; NAME/KEY: intron
; LOCATION: (1473)...(124390)
; NAME/KEY: exon
; LOCATION: (124391)...(124544)
; NAME/KEY: intron
; LOCATION: (124545)...(125409)
; NAME/KEY: exon
; LOCATION: (125410)...(125595)
; NAME/KEY: intron
; LOCATION: (125596)...(128711)
; NAME/KEY: exon
; LOCATION: (128712)...(128848)
; NAME/KEY: intron
; LOCATION: (128849)...(133400)
; NAME/KEY: exon
; LOCATION: (133401)...(133469)
; NAME/KEY: intron
; LOCATION: (133470)...(134652)
; NAME/KEY: exon
; LOCATION: (134653)...(134773)
; NAME/KEY: intron
; LOCATION: (134774)...(136116)
; NAME/KEY: exon
; LOCATION: (136117)...(136261)
; NAME/KEY: intron
; LOCATION: (136262)...(137936)
; NAME/KEY: exon
; LOCATION: (137937)...(138053)
; NAME/KEY: intron
; LOCATION: (138054)...(138637)
; NAME/KEY: exon
; LOCATION: (138638)...(138766)
; NAME/KEY: intron
; LOCATION: (138767)...(138864)
; NAME/KEY: exon
; LOCATION: (138865)...(138940)
; NAME/KEY: intron
; LOCATION: (138941)...(139765)
; NAME/KEY: exon
; LOCATION: (139766)...(139860)
; NAME/KEY: intron
; LOCATION: (139861)...(142245)
; NAME/KEY: exon
; LOCATION: (142246)...(142445)
; NAME/KEY: intron
; LOCATION: (142446)...(143605)
; NAME/KEY: exon
; LOCATION: (143606)...(143738)
; NAME/KEY: intron
; LOCATION: (143739)...(145838)
; NAME/KEY: exon
; LOCATION: (145839)...(145931)
; NAME/KEY: intron
; LOCATION: (145932)...(147385)

; NAME/KEY: exon
; LOCATION: (147386)...(147544)
; NAME/KEY: intron
; LOCATION: (147545)...(153274)
; NAME/KEY: exon
; LOCATION: (153275)...(153321)
; NAME/KEY: intron
; LOCATION: (153322)...(155088)
; NAME/KEY: exon
; LOCATION: (155089)...(155231)
; NAME/KEY: intron
; LOCATION: (155232)...(156025)
; NAME/KEY: exon
; LOCATION: (156026)...(156151)
; NAME/KEY: intron
; LOCATION: (156152)...(156826)
; NAME/KEY: exon
; LOCATION: (156827)...(156928)
; NAME/KEY: intron
; LOCATION: (156929)...(163399)
; NAME/KEY: exon
; LOCATION: (163400)...(163586)
US-09-676-610B-24

Query Match 60.0%; Score 18; DB 4; Length 169998;
Best Local Similarity 80.8%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GCAGACGAGGATGACCAATGAG 30
|||||
Db 107252 GCAGAGCTGGATGCCCACTGAG 107227

RESULT 19

US-09-877-177A-10/c
; Sequence 10, Application US/09877177A
; Patent No. 658219
; GENERAL INFORMATION:
; APPLICANT: K. Danenberg
; TITLE OF INVENTION: Method of determining Epidermal Growth
; TITLE OF INVENTION: Factor Receptor and HER2-Neu Gene Expression
; TITLE OF INVENTION: and Correlation of Levels Thereof With Survival
; FILE REFERENCE: 11220/120
; CURRENT APPLICATION NUMBER: US/09/877,177A
; CURRENT FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 197496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-877-177A-10

Query Match 60.0%; Score 18; DB 4; Length 197496;
Best Local Similarity 80.8%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GCAGACGAGGATGACCAATGAG 30
|||||
Db 115252 GCAGAGCTGGATGCCCACTGAG 115227

RESULT 20

US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506591
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:

ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Query Match 60.0%; Score 18; DB 4; Length 1830121;
Best Local Similarity 80.8%; Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 TGGCGCAGACCAAGAGTAGTACCA 26
Db 1027675 TGGTCAGCAGCAGTAGTACCA 1027700
RESULT 21
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Query Match 60.0%; Score 18; DB 4; Length 1830121;
Best Local Similarity 80.8%; Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 1 TGGCGCAGACCAAGAGTAGTACCA 26
Db 1027675 TGGTCAGCAGCAGTAGTACCA 1027700
RESULT 22
US-08-973-544-2
Sequence 2, Application US/08973544
Patent No. 6338950
GENERAL INFORMATION:
APPLICANT: WEISS, Elisabeth
TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIKAI DO, MARCELSTEIN, MURRAY & ORAM LLP
STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,544
FILING DATE: 18-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT EP 96/02663
FILING DATE: 20-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95109511.6
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95112201.9
FILING DATE: 03-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kitts, Monica Chin
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P8341-7073
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 312 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..312
US-08-973-544-2

Query Match 59.3%; Score 17.8; DB 4; Length 312;
Best Local Similarity 75.9%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGATGACCAATGAG 30
Db 239 GGCGCAGACAGAGAGGACCAAGGAG 267

RESULT 23
US-08-973-544-1
Sequence 1, Application US/08973544
Patent No. 6338950
GENERAL INFORMATION:
APPLICANT: WEISS, Elisabeth
TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIKAI, DO, MARCELSTEIN, MURRAY & ORAM LLP
STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,544
FILING DATE: 18-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT EP 96/02663
FILING DATE: 20-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95109511.6
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95112201.9
FILING DATE: 03-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kitts, Monica Chin
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P8341-7073
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5581 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 48..162
NAME/KEY: exon
LOCATION: 544..652
NAME/KEY: exon

LOCATION: 1044..1162
FEATURE:
NAME/KEY: exon
LOCATION: 1475..1567
FEATURE:
NAME/KEY: exon
LOCATION: 1775..1797
FEATURE:
NAME/KEY: exon
LOCATION: 2325..2709
US-08-973-544-1

Query Match 59.3%; Score 17.8; DB 4; Length 5581;
Best Local Similarity 75.9%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGATGACCAATGAG 30
Db 2428 GGCGCAGACAGAGAGGACCAAGGAG 2456

RESULT 24
US-08-311-174-4/c
Sequence 4, Application US/08311174
Patent No. 5556776
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, MAKOTO
TITLE OF INVENTION: SUCRASE GENE DERIVED FROM CORYNEFORM
TITLE OF INVENTION: BACTERIA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,174
FILING DATE: 23-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 046836/1992
FILING DATE: 04-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5556776man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-699-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6911 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-311-174-4

Query Match 59.3%; Score 17.8; DB 1; Length 6911;
Best Local Similarity 75.9%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGATGACCAATGAG 30

| | | | |
|--|------|--------------------------------|------|
| Db | 2963 | GGAGCAGAACGAGTTCGTGTCATCACTGAG | 2935 |
| RESULT 25 | | | |
| US-09-621-976-1469 | | | |
| ; Sequence 1469, Application US/09621976 | | | |
| ; Patent No. 6639063 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Dumas Milne Edwards, J.B. | | | |
| ; APPLICANT: Jobert, S. | | | |
| ; APPLICANT: Giordano, J.Y. | | | |
| ; TITLE OF INVENTION: ESTs and Encoded Human Proteins. | | | |
| ; FILE REFERENCE: GENSET.054PR2 | | | |
| ; CURRENT APPLICATION NUMBER: US/09/621.976 | | | |
| ; CURRENT FILING DATE: 2000-07-21 | | | |
| ; NUMBER OF SEQ ID NOS: 19335 | | | |
| ; SOFTWARE: Patent.pm | | | |
| ; SEQ ID NO 1469 | | | |
| ; LENGTH: 481 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Homo sapiens | | | |
| ; FEATURE: | | | |
| ; NAME/KEY: CDS | | | |
| ; LOCATION: 100..441 | | | |
| ; NAME/KEY: sig_peptide | | | |
| ; LOCATION: 100..303 | | | |
| ; OTHER INFORMATION: Von Heijne matrix | | | |
| ; OTHER INFORMATION: score 7.5 | | | |
| ; OTHER INFORMATION: seq VLMFSQVISICWA/AM | | | |
| US-09-621-976-1469 | | | |
| Query Match 58.7%; Score 17.6; DB 4; Length 481; | | | |
| Best Local Similarity 83.3%; Pred. No. 1.6e+02; | | | |
| Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0; | | | |
| Qy | 7 | AGAACCCAGGAGTACACCAATGAG | 30 |
| Db | 396 | AGTACCGGAGGAGCACCACTGAG | 419 |
| RESULT 26 | | | |
| US-09-184-418C-99 | | | |
| ; Sequence 99, Application US/09184418C | | | |
| ; Patent No. 6492110 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Hahn, Beatrice | | | |
| ; APPLICANT: Gao, Feng | | | |
| ; APPLICANT: Shaw, George | | | |
| ; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN | | | |
| ; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1 | | | |
| ; FILE REFERENCE: D6287 | | | |
| ; CURRENT APPLICATION NUMBER: US/09/184.418C | | | |
| ; CURRENT FILING DATE: 1999-11-02 | | | |
| ; NUMBER OF SEQ ID NOS: 112 | | | |
| ; SEQ ID NO 99 | | | |
| ; LENGTH: 654 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Human immunodeficiency virus type 1 | | | |
| ; FEATURE: | | | |
| ; OTHER INFORMATION: isolate-94CV017.41; genes=nef | | | |
| US-09-184-418C-99 | | | |
| Query Match 58.7%; Score 17.6; DB 4; Length 654; | | | |
| Best Local Similarity 83.3%; Pred. No. 1.7e+02; | | | |
| Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0; | | | |
| Qy | 5 | GCAGAACCCAGGAGTACCAATG | 28 |
| Db | 106 | GCAGCACGAGGATGAGGACGATG | 129 |
| RESULT 27 | | | |
| US-09-620-312D-29 | | | |
| ; Sequence 29, Application US/09620312D | | | |
| ; Patent No. 6569662 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Tang, Y. Tom | | | |
| ; APPLICANT: Liu, Chenghua | | | |
| ; APPLICANT: Asundi, Vinod | | | |
| ; APPLICANT: Zhang, Jie | | | |
| ; APPLICANT: Ren, Feiyan | | | |
| ; APPLICANT: Chen, Rui-hong | | | |
| ; APPLICANT: Zhao, Qing A. | | | |
| ; APPLICANT: Wehrman, Tom | | | |
| ; APPLICANT: Xue, Aidong J. | | | |
| ; APPLICANT: Yang, Yonghong | | | |
| ; APPLICANT: Wang, Jian-Rui | | | |
| ; APPLICANT: Zhou, Ping | | | |
| ; APPLICANT: Ma, Yunding | | | |
| ; APPLICANT: Wang, Dunrui | | | |
| ; APPLICANT: Wang, Zhiwei | | | |
| ; APPLICANT: John Tillinghast | | | |
| ; APPLICANT: Drmanac, Radcoje T. | | | |
| ; TITLE OF INVENTION: No. 6569662el Nucleic Acids and | | | |
| ; TITLE OF INVENTION: Polypeptides | | | |
| ; FILE REFERENCE: 784CIP2B | | | |
| ; CURRENT APPLICATION NUMBER: US/09/620.312D | | | |
| ; CURRENT FILING DATE: 2000-07-19 | | | |
| ; PRIOR APPLICATION NUMBER: 09/552,317 | | | |
| ; PRIOR FILING DATE: 2000-04-25 | | | |
| ; PRIOR APPLICATION NUMBER: 09/488,725 | | | |
| ; PRIOR FILING DATE: 2000-01-21 | | | |
| ; NUMBER OF SEQ ID NOS: 1105 | | | |
| ; SOFTWARE: pt_FL_genes Version 1.0 | | | |
| ; SEQ ID NO 29 | | | |
| ; LENGTH: 1813 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Homo sapiens | | | |
| ; FEATURE: | | | |
| ; NAME/KEY: CDS | | | |
| ; LOCATION: (221)..(1189) | | | |
| US-09-620-312D-29 | | | |
| Query Match 58.7%; Score 17.6; DB 4; Length 1813; | | | |
| Best Local Similarity 83.3%; Pred. No. 2.1e+02; | | | |
| Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0; | | | |

Db 767 GAGCAGGACCAGTAGGAGCAGAAATGA 741

Search completed: June 20, 2004, 11:47:23
JOB time : 50.2996 secsQY 5 GCAGAACGAGGAGTAGCACCACCAATG 28
Db 258 GCAGTAGCAGCAGCAGCACCACCAATG 281

RESULT 29
US-09-184-418C-10
; Sequence 10, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184,418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 10
; LENGTH: 9060
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: Isolates=94CY017.41; 159.1649:"gag";
; OTHER INFORMATION: 1442.4453:"pol"; 4398.4976:"vif"; 4916.5206:"vpr";
; OTHER INFORMATION: 5187.7841:"tat"; 5326.8046:"rev";
; OTHER INFORMATION: 5428.5673:"vpu"; 5591.8188:"env"; 8190.8843:"nsf"
US-09-184-418C-10

Query Match 58.7%; Score 17.6; DB 4; Length 9060;
Best Local Similarity 83.3%; Pred.No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCAGAACGAGGAGTAGCACCACCAATG 28
Db 8295 GCAGCAGGAGGAGTAGGAGCAGATG 8318

RESULT 30
US-09-598-401C-107/c
; Sequence 107, Application US/09598401C
; Patent No. 6596925
; GENERAL INFORMATION:
; APPLICANT: Perera, J. Ranjan
; APPLICANT: Eagleton, Claire
; APPLICANT: Rice, Stephen J.
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Expression
; FILE REFERENCE: 11000.1036c2
; CURRENT APPLICATION NUMBER: US/09/598,401C
; CURRENT FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: U.S. No. 6596925 60/146,591
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: PCT/NZ00/00018
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: U.S. No. 6596925 09/276,599
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-598-401C-107

Query Match 58.0%; Score 17.4; DB 4; Length 948;
Best Local Similarity 77.8%; Pred.No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GCGCAGAACGAGGAGTAGCACCACCAATGA 29

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 159.683 Seconds

(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggccagaaacaggagtagcaccatgag 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 30 | 100.0 | 1584 | 9 | US-09-712-363-10 |
| 2 | 20.6 | 68.7 | 176 | 15 | US-10-029-386-25513 |
| 3 | 20.6 | 68.7 | 501 | 15 | US-10-029-386-11813 |
| 4 | 20.6 | 68.7 | 3045 | 9 | US-09-954-456-218 |
| 5 | 20.6 | 68.7 | 3045 | 9 | US-09-880-107-2319 |
| 6 | 20.6 | 68.7 | 3045 | 13 | US-10-342-887-400 |
| 7 | 20.6 | 68.7 | 3045 | 13 | US-10-172-118-400 |
| 8 | 20.6 | 68.7 | 3050 | 13 | US-10-342-887-525 |
| 9 | 20.6 | 68.7 | 3050 | 13 | US-10-172-118-525 |
| 10 | 20.6 | 68.7 | 42547 | 15 | US-10-268-822-12 |
| 11 | 19.6 | 65.3 | 629 | 13 | US-10-027-632-22664 |
| 12 | 19.6 | 65.3 | 629 | 16 | US-10-027-632-22664 |
| 13 | 19.6 | 65.3 | 99934 | 16 | US-10-085-117-73 |
| 14 | 19.4 | 64.7 | 360 | 16 | US-10-062-674-960 |

| | | | | | | |
|----|------|------|---------|----|----------------------|--------------------|
| 15 | 19.4 | 64.7 | 537 | 9 | US-09-864-761-7559 | Sequence 7559, Ap |
| 16 | 19.4 | 64.7 | 1260 | 16 | US-10-371-264-15 | Sequence 15, Appl |
| 17 | 19.4 | 64.7 | 1260 | 16 | US-10-371-099-15 | Sequence 15, Appl |
| 18 | 19.4 | 64.7 | 1260 | 16 | US-10-371-122-15 | Sequence 15, Appl |
| 19 | 19.4 | 64.7 | 1260 | 16 | US-10-373-567-15 | Sequence 15, Appl |
| 20 | 19.4 | 64.7 | 1260 | 17 | US-10-628-088-15 | Sequence 15, Appl |
| 21 | 19.4 | 64.7 | 1779 | 16 | US-10-369-493-25221 | Sequence 25221, A |
| 22 | 19.2 | 64.0 | 1342 | 13 | US-10-425-114-15923 | Sequence 15923, A |
| 23 | 19.2 | 64.0 | 1950 | 16 | US-10-369-493-46459 | Sequence 46459, A |
| 24 | 19.2 | 64.0 | 21706 | 13 | US-10-158-844-36 | Sequence 36, Appl |
| 25 | 19 | 63.3 | 393 | 10 | US-09-918-998-5024 | Sequence 5024, Ap |
| 26 | 19 | 63.3 | 637 | 9 | US-09-563-817-317 | Sequence 317, App |
| 27 | 19 | 63.3 | 843 | 13 | US-10-092-900A-119 | Sequence 119, App |
| 28 | 19 | 63.3 | 40178 | 13 | US-10-282-174-467 | Sequence 467, App |
| 29 | 19 | 63.3 | 40178 | 13 | US-10-282-174-468 | Sequence 468, App |
| 30 | 19 | 63.3 | 495269 | 16 | US-10-398-221-8 | Sequence 8, Appl |
| 31 | 19 | 63.3 | 3011208 | 16 | US-10-398-221-8 | Sequence 2058, Ap |
| 32 | 18.8 | 62.7 | 343 | 13 | US-10-027-632-177235 | Sequence 177235, A |
| 33 | 18.8 | 62.7 | 343 | 16 | US-10-027-632-177235 | Sequence 177235, A |
| 34 | 18.8 | 62.7 | 451 | 15 | US-10-198-846-4908 | Sequence 4908, Ap |
| 35 | 18.8 | 62.7 | 574 | 15 | US-10-198-846-10080 | Sequence 10080, A |
| 36 | 18.8 | 62.7 | 636 | 15 | US-10-198-846-4977 | Sequence 4977, Ap |
| 37 | 18.8 | 62.7 | 646 | 15 | US-10-198-846-9104 | Sequence 9104, Ap |
| 38 | 18.8 | 62.7 | 663 | 13 | US-10-027-632-231136 | Sequence 231136, A |
| 39 | 18.8 | 62.7 | 663 | 16 | US-10-027-632-231136 | Sequence 231136, A |
| 40 | 18.8 | 62.7 | 865 | 15 | US-10-198-846-4870 | Sequence 4870, Ap |
| 41 | 18.8 | 62.7 | 868 | 15 | US-10-198-846-5195 | Sequence 5195, Ap |
| 42 | 18.8 | 62.7 | 1143 | 9 | US-09-006-298-24 | Sequence 24, Appl |
| 43 | 18.8 | 62.7 | 1440 | 14 | US-10-044-090-820 | Sequence 820, App |
| 44 | 18.8 | 62.7 | 2331 | 16 | US-10-369-493-35883 | Sequence 35883, A |
| 45 | 18.8 | 62.7 | 2907 | 9 | US-09-954-456-518 | Sequence 518, App |
| 46 | 18.8 | 62.7 | 2907 | 9 | US-09-954-456-823 | Sequence 823, App |
| 47 | 18.8 | 62.7 | 2907 | 9 | US-09-954-456-1226 | Sequence 1226, Ap |
| 48 | 18.8 | 62.7 | 2907 | 9 | US-09-880-107-2318 | Sequence 2318, Ap |
| 49 | 18.8 | 62.7 | 2907 | 13 | US-10-342-887-399 | Sequence 399, App |
| 50 | 18.8 | 62.7 | 2907 | 13 | US-10-172-118-399 | Sequence 399, App |
| 51 | 18.8 | 62.7 | 2907 | 17 | US-10-641-643-1053 | Sequence 1053, Ap |
| 52 | 18.6 | 62.0 | 584 | 13 | US-10-027-632-255998 | Sequence 255998, A |
| 53 | 18.6 | 62.0 | 584 | 16 | US-10-027-632-255998 | Sequence 255998, A |
| 54 | 18.6 | 62.0 | 996 | 13 | US-10-440-503-219 | Sequence 219, App |
| 55 | 18.6 | 62.0 | 996 | 15 | US-10-461-925-219 | Sequence 219, App |
| 56 | 18.6 | 62.0 | 996 | 15 | US-10-146-772-219 | Sequence 219, App |
| 57 | 18.6 | 62.0 | 996 | 16 | US-10-241-742-219 | Sequence 219, App |
| 58 | 18.6 | 62.0 | 996 | 16 | US-10-440-523-219 | Sequence 219, App |
| 59 | 18.6 | 62.0 | 1698 | 13 | US-10-425-114-1399 | Sequence 1399, App |
| 60 | 18.6 | 62.0 | 1701 | 9 | US-09-801-368-417 | Sequence 417, App |
| 61 | 18.6 | 62.0 | 2369 | 16 | US-10-108-260A-1248 | Sequence 1248, Ap |
| 62 | 18.6 | 62.0 | 104083 | 13 | US-10-087-192-670 | Sequence 732, App |
| 63 | 18.4 | 61.3 | 546 | 9 | US-09-919-580-732 | Sequence 732, App |
| 64 | 18.4 | 61.3 | 634 | 13 | US-10-424-599-75105 | Sequence 75105, A |
| 65 | 18.4 | 61.3 | 729 | 9 | US-09-764-855-160 | Sequence 160, App |
| 66 | 18.4 | 61.3 | 729 | 15 | US-10-072-349-160 | Sequence 160, App |
| 67 | 18.4 | 61.3 | 742 | 13 | US-10-027-632-141364 | Sequence 141364, A |
| 68 | 18.4 | 61.3 | 742 | 16 | US-10-027-632-141364 | Sequence 141364, A |
| 69 | 18.4 | 61.3 | 752 | 16 | US-10-264-049-2070 | Sequence 2070, Ap |
| 70 | 18.4 | 61.3 | 778 | 15 | US-10-106-698-589 | Sequence 589, App |
| 71 | 18.4 | 61.3 | 813 | 13 | US-10-027-632-160048 | Sequence 160048, A |
| 72 | 18.4 | 61.3 | 813 | 16 | US-10-027-632-160048 | Sequence 160048, A |
| 73 | 18.4 | 61.3 | 864 | 9 | US-09-070-927A-767 | Sequence 767, App |
| 74 | 18.4 | 61.3 | 888 | 9 | US-09-764-855-159 | Sequence 159, App |
| 75 | 18.4 | 61.3 | 888 | 15 | US-10-072-349-159 | Sequence 159, App |
| 76 | 18.4 | 61.3 | 1059 | 9 | US-09-938-842A-1816 | Sequence 1816, Ap |
| 77 | 18.4 | 61.3 | 1059 | 11 | US-09-938-842A-1816 | Sequence 1816, Ap |
| 78 | 18.4 | 61.3 | 1246 | 10 | US-09-934-455-345 | Sequence 345, App |
| 79 | 18.4 | 61.3 | 1246 | 13 | US-10-412-699B-13 | Sequence 13, Appl |
| 80 | 18.4 | 61.3 | 1246 | 13 | US-10-412-699B-1713 | Sequence 1713, App |
| 81 | 18.4 | 61.3 | 1246 | 13 | US-10-225-066A-587 | Sequence 587, App |
| 82 | 18.4 | 61.3 | 1246 | 15 | US-10-278-173-123 | Sequence 123, App |
| 83 | 18.4 | 61.3 | 1246 | 15 | US-10-295-403-7 | Sequence 7, Appl |
| 84 | 18.4 | 61.3 | 1246 | 15 | US-10-278-536-171 | Sequence 171, App |
| 85 | 18.4 | 61.3 | 1246 | 16 | US-10-374-780A-1949 | Sequence 1949, App |
| 86 | 18.4 | 61.3 | 1246 | 17 | US-10-685-522-3 | Sequence 3, Appl |
| 87 | 18.4 | 61.3 | 1315 | 16 | US-10-310-154-235 | Sequence 235, App |

| | | | | | | | | | | | | | |
|-------|------|------|---------|----|----------------------|--------------------|-------|------|------|---------|----|----------------------|---------------------|
| C 88 | 18.4 | 61.3 | 2471 | 16 | US-10-367-980A-4 | Sequence 4, Appli | C 161 | 17.8 | 59.3 | 866 | 15 | US-10-198-846-4450 | Sequence 4450, Ap |
| C 89 | 18.4 | 61.3 | 2855 | 10 | US-09-822-846-155 | Sequence 155, Appl | C 162 | 17.8 | 59.3 | 946 | 12 | US-09-876-143-1726 | Sequence 1726, Ap |
| C 90 | 18.4 | 61.3 | 7058 | 15 | US-10-287-218-38 | Sequence 38, Appl | C 163 | 17.8 | 59.3 | 1286 | 13 | US-10-027-632-230147 | Sequence 230147, Ap |
| C 91 | 18.4 | 61.3 | 7433 | 15 | US-10-198-846-13220 | Sequence 13220, A | C 164 | 17.8 | 59.3 | 1286 | 16 | US-10-027-632-230147 | Sequence 230147, Ap |
| C 92 | 18.2 | 60.7 | 848 | 15 | US-10-029-386-24158 | Sequence 24158, A | C 165 | 17.8 | 59.3 | 1299 | 9 | US-09-738-626-2305 | Sequence 2305, Ap |
| C 93 | 18.2 | 60.7 | 7931 | 16 | US-10-367-978-31 | Sequence 31, Appl | C 166 | 17.8 | 59.3 | 1886 | 13 | US-10-020-120-31 | Sequence 31, Appl |
| C 94 | 18.2 | 60.7 | 14955 | 16 | US-10-144-198-29 | Sequence 29, Appl | C 167 | 17.8 | 59.3 | 2000 | 16 | US-10-260-238-2369 | Sequence 2369, Ap |
| C 95 | 18 | 60.0 | 239 | 13 | US-10-424-599-1079 | Sequence 1079, Ap | C 168 | 17.8 | 59.3 | 2303 | 15 | US-10-380-558-5 | Sequence 5, Appli |
| C 96 | 18 | 60.0 | 406 | 13 | US-09-878-722-124 | Sequence 124, Ap | C 169 | 17.8 | 59.3 | 2477 | 13 | US-10-027-632-110037 | Sequence 110037, Ap |
| C 97 | 18 | 60.0 | 406 | 13 | US-09-878-722-124 | Sequence 124, Ap | C 170 | 17.8 | 59.3 | 2477 | 13 | US-10-027-632-110038 | Sequence 110038, Ap |
| C 98 | 18 | 60.0 | 413 | 9 | US-09-960-352-7890 | Sequence 7890, Ap | C 171 | 17.8 | 59.3 | 2477 | 13 | US-10-027-632-110039 | Sequence 110039, Ap |
| C 99 | 18 | 60.0 | 424 | 9 | US-09-960-352-308 | Sequence 308, Ap | C 172 | 17.8 | 59.3 | 2477 | 16 | US-10-027-632-110037 | Sequence 110037, Ap |
| C 100 | 18 | 60.0 | 427 | 9 | US-09-864-761-89 | Sequence 89, Appl | C 173 | 17.8 | 59.3 | 2477 | 16 | US-10-027-632-110038 | Sequence 110038, Ap |
| C 101 | 18 | 60.0 | 468 | 9 | US-09-815-242-7062 | Sequence 7062, Ap | C 174 | 17.8 | 59.3 | 2477 | 16 | US-10-027-632-110039 | Sequence 110039, Ap |
| C 102 | 18 | 60.0 | 468 | 13 | US-10-282-122A-22184 | Sequence 22184, A | C 175 | 17.8 | 59.3 | 3125 | 14 | US-10-044-090-16 | Sequence 16, Appl |
| C 103 | 18 | 60.0 | 635 | 13 | US-10-027-632-14842 | Sequence 14842, A | C 176 | 17.8 | 59.3 | 113604 | 13 | US-10-027-152B-1 | Sequence 1, Appli |
| C 104 | 18 | 60.0 | 635 | 16 | US-10-027-632-14842 | Sequence 14842, A | C 177 | 17.8 | 59.3 | 113604 | 13 | US-10-027-152B-2 | Sequence 2, Appli |
| C 105 | 18 | 60.0 | 672 | 9 | US-09-864-761-16927 | Sequence 16927, A | C 178 | 17.8 | 59.3 | 113604 | 15 | US-10-027-195A-1 | Sequence 1, Appli |
| C 106 | 18 | 60.0 | 824 | 9 | US-09-770-445-716 | Sequence 716, Ap | C 179 | 17.8 | 59.3 | 113604 | 15 | US-10-027-195A-2 | Sequence 2, Appli |
| C 107 | 18 | 60.0 | 1062 | 9 | US-09-938-842A-2094 | Sequence 2094, Ap | C 180 | 17.8 | 59.3 | 251364 | 15 | US-10-175-523-58 | Sequence 58, Appl |
| C 108 | 18 | 60.0 | 1062 | 11 | US-09-938-842A-2094 | Sequence 2094, Ap | C 181 | 17.8 | 59.3 | 251364 | 15 | US-10-175-523-61 | Sequence 61, Appl |
| C 109 | 18 | 60.0 | 1136 | 9 | US-09-917-800A-1671 | Sequence 1671, Ap | C 182 | 17.8 | 59.3 | 251364 | 15 | US-10-175-523-79 | Sequence 79, Appl |
| C 110 | 18 | 60.0 | 1136 | 12 | US-10-152-218A-1722 | Sequence 1722, Ap | C 183 | 17.8 | 59.3 | 3309400 | 9 | US-09-738-626-1 | Sequence 1, Appli |
| C 111 | 18 | 60.0 | 1136 | 16 | US-10-191-803-121 | Sequence 121, Ap | C 184 | 17.6 | 58.7 | 359 | 9 | US-09-728-445-651 | Sequence 651, Appl |
| C 112 | 18 | 60.0 | 1267 | 9 | US-09-770-445-25 | Sequence 25, Appl | C 185 | 17.6 | 58.7 | 393 | 13 | US-10-424-599-89079 | Sequence 89079, A |
| C 113 | 18 | 60.0 | 1932 | 15 | US-10-022-390-660 | Sequence 660, Ap | C 186 | 17.6 | 58.7 | 400 | 9 | US-09-789-919-18 | Sequence 18, Appl |
| C 114 | 18 | 60.0 | 1932 | 15 | US-10-022-390-667 | Sequence 667, Ap | C 187 | 17.6 | 58.7 | 410 | 10 | US-09-803-719-1803 | Sequence 1803, Ap |
| C 115 | 18 | 60.0 | 1932 | 15 | US-10-022-390-688 | Sequence 688, Ap | C 188 | 17.6 | 58.7 | 418 | 13 | US-10-424-599-88127 | Sequence 88127, A |
| C 116 | 18 | 60.0 | 1932 | 15 | US-10-022-390-691 | Sequence 691, Ap | C 189 | 17.6 | 58.7 | 503 | 16 | US-10-341-961A-84 | Sequence 84, Appl |
| C 117 | 18 | 60.0 | 1932 | 15 | US-10-022-390-693 | Sequence 693, Ap | C 190 | 17.6 | 58.7 | 514 | 9 | US-09-731-872-153 | Sequence 153, Ap |
| C 118 | 18 | 60.0 | 1932 | 15 | US-10-022-390-700 | Sequence 700, Ap | C 191 | 17.6 | 58.7 | 514 | 10 | US-09-876-597-153 | Sequence 153, Ap |
| C 119 | 18 | 60.0 | 1932 | 15 | US-10-022-390-702 | Sequence 702, Ap | C 192 | 17.6 | 58.7 | 588 | 15 | US-10-029-386-2780 | Sequence 2780, Ap |
| C 120 | 18 | 60.0 | 1932 | 15 | US-10-022-249-660 | Sequence 660, Ap | C 193 | 17.6 | 58.7 | 631 | 9 | US-09-764-870-130 | Sequence 130, Ap |
| C 121 | 18 | 60.0 | 1932 | 15 | US-10-022-249-667 | Sequence 667, Ap | C 194 | 17.6 | 58.7 | 631 | 11 | US-09-764-875-346 | Sequence 346, Ap |
| C 122 | 18 | 60.0 | 1932 | 15 | US-10-022-249-688 | Sequence 688, Ap | C 195 | 17.6 | 58.7 | 631 | 11 | US-10-125-540-130 | Sequence 130, Ap |
| C 123 | 18 | 60.0 | 1932 | 15 | US-10-022-249-691 | Sequence 691, Ap | C 196 | 17.6 | 58.7 | 654 | 15 | US-10-290-579-99 | Sequence 99, Appl |
| C 124 | 18 | 60.0 | 1932 | 15 | US-10-022-249-693 | Sequence 693, Ap | C 197 | 17.6 | 58.7 | 685 | 13 | US-10-424-599-6185 | Sequence 6185, Ap |
| C 125 | 18 | 60.0 | 1932 | 15 | US-10-022-249-700 | Sequence 700, Ap | C 198 | 17.6 | 58.7 | 929 | 17 | US-10-240-240A-26 | Sequence 26, Appl |
| C 126 | 18 | 60.0 | 1932 | 15 | US-10-022-249-702 | Sequence 702, Ap | C 199 | 17.6 | 58.7 | 1448 | 15 | US-10-197-666A-15 | Sequence 15, Appl |
| C 127 | 18 | 60.0 | 2166 | 15 | US-10-101-510-326 | Sequence 326, Ap | C 200 | 17.6 | 58.7 | 1448 | 15 | US-10-024-298A-38 | Sequence 38, Appl |
| C 128 | 18 | 60.0 | 4055 | 15 | US-10-233-045-9 | Sequence 9, Appli | C 201 | 17.6 | 58.7 | 1448 | 15 | US-10-042-211A-38 | Sequence 38, Appl |
| C 129 | 18 | 60.0 | 4716 | 15 | US-10-084-817-237 | Sequence 237, Ap | C 202 | 17.6 | 58.7 | 1448 | 17 | US-10-617-217A-38 | Sequence 38, Appl |
| C 130 | 18 | 60.0 | 87687 | 13 | US-10-087-192-1978 | Sequence 1978, Ap | C 203 | 17.6 | 58.7 | 1622 | 15 | US-10-029-386-22666 | Sequence 22666, A |
| C 131 | 18 | 60.0 | 189998 | 16 | US-10-380-931-24 | Sequence 24, Appl | C 204 | 17.6 | 58.7 | 1813 | 15 | US-10-037-270-29 | Sequence 29, Appl |
| C 132 | 18 | 60.0 | 197496 | 9 | US-09-877-177-10 | Sequence 10, Appl | C 205 | 17.6 | 58.7 | 1813 | 16 | US-10-117-722-29 | Sequence 29, Appl |
| C 133 | 18 | 60.0 | 197496 | 13 | US-10-426-836-10 | Sequence 10, Appl | C 206 | 17.6 | 58.7 | 2000 | 9 | US-09-887-576-837 | Sequence 837, Ap |
| C 134 | 18 | 60.0 | 493631 | 13 | US-10-087-192-205 | Sequence 205, Ap | C 207 | 17.6 | 58.7 | 2000 | 16 | US-10-260-238-1733 | Sequence 1733, Ap |
| C 135 | 18 | 60.0 | 1830121 | 15 | US-10-329-960-1 | Sequence 1, Appli | C 208 | 17.6 | 58.7 | 2167 | 9 | US-09-965-529-46 | Sequence 46, Appl |
| C 136 | 18 | 60.0 | 1830121 | 16 | US-10-329-960-1 | Sequence 1, Appli | C 209 | 17.6 | 58.7 | 2167 | 10 | US-09-969-680A-46 | Sequence 46, Appl |
| C 137 | 17.8 | 59.3 | 285 | 9 | US-09-294-093B-2124 | Sequence 2124, Ap | C 210 | 17.6 | 58.7 | 2167 | 10 | US-10-264-049-784 | Sequence 784, Ap |
| C 138 | 17.8 | 59.3 | 290 | 9 | US-09-294-093B-829 | Sequence 829, Ap | C 211 | 17.6 | 58.7 | 2343 | 16 | US-10-282-122A-35133 | Sequence 35133, A |
| C 139 | 17.8 | 59.3 | 309 | 13 | US-10-424-599-96694 | Sequence 96694, A | C 212 | 17.6 | 58.7 | 3530 | 9 | US-09-784-847-1985 | Sequence 1985, Ap |
| C 140 | 17.8 | 59.3 | 321 | 9 | US-09-960-352-6356 | Sequence 6356, Ap | C 213 | 17.6 | 58.7 | 3530 | 15 | US-10-092-154-1495 | Sequence 1495, Ap |
| C 141 | 17.8 | 59.3 | 353 | 9 | US-09-960-352-451 | Sequence 451, Ap | C 214 | 17.6 | 58.7 | 3530 | 15 | US-10-087-192-586 | Sequence 586, Ap |
| C 142 | 17.8 | 59.3 | 379 | 13 | US-10-425-114-25635 | Sequence 25635, A | C 215 | 17.6 | 58.7 | 3536 | 15 | US-10-092-154-1494 | Sequence 1494, Ap |
| C 143 | 17.8 | 59.3 | 426 | 15 | US-10-198-846-4939 | Sequence 4939, Ap | C 216 | 17.6 | 58.7 | 5241 | 16 | US-10-260-238-1482 | Sequence 1482, Ap |
| C 144 | 17.8 | 59.3 | 435 | 9 | US-09-880-107-2653 | Sequence 2653, Ap | C 217 | 17.6 | 58.7 | 5241 | 16 | US-10-260-238-1482 | Sequence 1482, Ap |
| C 145 | 17.8 | 59.3 | 450 | 15 | US-10-240-965-182 | Sequence 182, Ap | C 218 | 17.6 | 58.7 | 5460 | 13 | US-10-282-122A-15354 | Sequence 15354, A |
| C 146 | 17.8 | 59.3 | 472 | 13 | US-10-085-783A-55343 | Sequence 55343, A | C 219 | 17.6 | 58.7 | 9060 | 15 | US-10-290-579-10 | Sequence 10, Appl |
| C 147 | 17.8 | 59.3 | 472 | 16 | US-10-242-535A-55343 | Sequence 55343, A | C 220 | 17.6 | 58.7 | 16825 | 9 | US-09-784-847-1495 | Sequence 1495, Ap |
| C 148 | 17.8 | 59.3 | 550 | 15 | US-10-198-846-12307 | Sequence 12307, A | C 221 | 17.6 | 58.7 | 16825 | 15 | US-10-092-154-1495 | Sequence 1495, Ap |
| C 149 | 17.8 | 59.3 | 568 | 13 | US-10-027-632-22225 | Sequence 22225, A | C 222 | 17.4 | 58.0 | 393 | 15 | US-10-029-386-24571 | Sequence 24571, A |
| C 150 | 17.8 | 59.3 | 568 | 13 | US-10-027-632-22226 | Sequence 22226, A | C 223 | 17.4 | 58.0 | 455 | 13 | US-10-027-632-193088 | Sequence 193088, A |
| C 151 | 17.8 | 59.3 | 568 | 16 | US-10-027-632-22226 | Sequence 22226, A | C 224 | 17.4 | 58.0 | 455 | 13 | US-10-027-632-193088 | Sequence 193088, A |
| C 152 | 17.8 | 59.3 | 568 | 16 | US-10-027-632-22226 | Sequence 22226, A | C 225 | 17.4 | 58.0 | 455 | 16 | US-10-027-632-193088 | Sequence 193088, A |
| C 153 | 17.8 | 59.3 | 584 | 13 | US-10-027-632-186339 | Sequence 186339, A | C 226 | 17.4 | 58.0 | 483 | 13 | US-10-027-632-193089 | Sequence 193089, A |
| C 154 | 17.8 | 59.3 | 584 | 13 | US-10-027-632-186340 | Sequence 186340, A | C 227 | 17.4 | 58.0 | 483 | 13 | US-10-424-599-124170 | Sequence 124170, A |
| C 155 | 17.8 | 59.3 | 584 | 13 | US-10-221-945-8 | Sequence 8, Appli | C 228 | 17.4 | 58.0 | 550 | 15 | US-10-029-386-10868 | Sequence 10868, A |
| C 156 | 17.8 | 59.3 | 584 | 16 | US-10-027-632-186339 | Sequence 186339, A | C 229 | 17.4 | 58.0 | 638 | 13 | US-10-424-599-5220 | Sequence 5220, Ap |
| C 157 | 17.8 | 59.3 | 584 | 16 | US-10-027-632-186340 | Sequence 186340, A | C 230 | 17.4 | 58.0 | 688 | 13 | US-10-425-114-26025 | Sequence 26025, A |
| C 158 | 17.8 | 59.3 | 625 | 13 | US-09-973-278-697 | Sequence 697, Ap | C 231 | 17.4 | 58.0 | 712 | 13 | US-10-092-900A-117 | Sequence 117, Appl |
| C 159 | 17.8 | 59.3 | 633 | 13 | US-10-424-599-95348 | Sequence 95348, A | C 232 | 17.4 | 58.0 | 776 | 13 | US-10-027-632-104125 | Sequence 104125, A |
| C 160 | 17.8 | 59.3 | 678 | 15 | US-10-198-846-1499 | Sequence 1499, Ap | C 233 | 17.4 | 58.0 | 776 | 13 | US-10-027-632-325003 | Sequence 325003, A |

234 17.4 58.0 776 16 US-10-027-632-104125
235 17.4 58.0 776 16 US-10-027-632-325003
C 236 17.4 58.0 798 13 US-10-425-114-11559
C 237 17.4 58.0 904 15 US-10-198-846-3589
C 238 17.4 58.0 948 15 US-10-137-036-107
C 239 17.4 58.0 1314 10 US-09-968-433-48
C 240 17.4 58.0 1360 13 US-10-424-599-124168
C 241 17.4 58.0 1364 13 US-10-424-599-87503
C 242 17.4 58.0 1710 13 US-10-424-599-139445
C 243 17.4 58.0 1766 15 US-10-137-036-117
C 244 17.4 58.0 1981 13 US-10-425-114-2715
C 245 17.4 58.0 2008 13 US-10-425-114-21189
C 246 17.4 58.0 2068 16 US-10-104-047-63
C 247 17.4 58.0 2164 13 US-10-027-632-250839
C 248 17.4 58.0 2164 16 US-10-027-632-250839
C 249 17.4 58.0 2396 13 US-10-425-114-11406
C 250 17.4 58.0 2511 16 US-10-242-355-1174
C 251 17.4 58.0 2512 16 US-10-104-047-1888
C 252 17.4 58.0 2799 13 US-10-424-599-98143
C 253 17.4 58.0 3708 13 US-10-282-122A-24775
C 254 17.4 58.0 3708 13 US-10-170-385-416
C 255 17.4 58.0 3939 15 US-10-101-510-113
C 256 17.4 58.0 4464 9 US-09-738-626-1995
C 257 17.4 58.0 7158 9 US-09-974-300-2171
C 258 17.4 58.0 165961 16 US-10-085-117-241
C 259 17.4 58.0 194945 16 US-10-085-117-355
C 260 17.2 57.3 60 10 US-09-908-975-19254
C 261 17.2 57.3 214 9 US-09-864-761-30337
C 262 17.2 57.3 283 9 US-09-294-093B-5662
C 263 17.2 57.3 308 13 US-10-424-599-43021
C 264 17.2 57.3 315 13 US-10-424-599-20372
C 265 17.2 57.3 381 13 US-10-424-599-16271
C 266 17.2 57.3 404 16 US-10-027-632-45528
C 267 17.2 57.3 419 16 US-10-369-493-34763
C 268 17.2 57.3 419 13 US-10-027-632-309283
C 269 17.2 57.3 419 16 US-10-027-632-309283
C 270 17.2 57.3 438 15 US-10-198-846-12746
C 271 17.2 57.3 459 9 US-09-864-761-11125
C 272 17.2 57.3 470 13 US-10-027-632-35309
C 273 17.2 57.3 470 16 US-10-027-632-35309
C 274 17.2 57.3 478 10 US-09-918-995-26453
C 275 17.2 57.3 489 13 US-10-335-977-2911
C 276 17.2 57.3 500 10 US-09-918-995-31709
C 277 17.2 57.3 540 13 US-10-335-977-2912
C 278 17.2 57.3 546 13 US-10-027-632-5002
C 279 17.2 57.3 552 16 US-10-027-632-5002
C 280 17.2 57.3 552 13 US-10-424-599-111909
C 281 17.2 57.3 572 13 US-10-027-632-217089
C 282 17.2 57.3 572 16 US-10-027-632-217089
C 283 17.2 57.3 602 13 US-10-027-632-273182
C 284 17.2 57.3 602 16 US-10-027-632-273182
C 285 17.2 57.3 611 9 US-09-919-580-277
C 286 17.2 57.3 634 13 US-10-027-632-62359
C 287 17.2 57.3 634 13 US-10-027-632-297673
C 288 17.2 57.3 634 16 US-10-027-632-62359
C 289 17.2 57.3 634 16 US-10-027-632-297673
C 290 17.2 57.3 639 13 US-10-424-599-135447
C 291 17.2 57.3 807 13 US-10-027-632-165906
C 292 17.2 57.3 807 13 US-10-027-632-165906
C 293 17.2 57.3 807 13 US-10-027-632-165908
C 294 17.2 57.3 807 16 US-10-027-632-165908
C 295 17.2 57.3 807 16 US-10-027-632-165907
C 296 17.2 57.3 807 16 US-10-027-632-165908
C 297 17.2 57.3 846 16 US-10-291-265-660
C 298 17.2 57.3 902 16 US-10-291-265-188
C 299 17.2 57.3 963 16 US-10-369-493-43575
C 300 17.2 57.3 1041 13 US-10-027-632-248911

ALIGNMENTS

US-09-712-363-10
; Sequence 10, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/136,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-10

Query Match 100.0%; Score 30; DB 9; Length 1584;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCAGACCCAGGAGTAGCACCACATGAG 30
DB 441 TGGCGCAGACCCAGGAGTAGCACCACATGAG 470

RESULT 2

US-10-029-386-25513
; Sequence 25513, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25513
; LENGTH: 176
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR19.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: gi14572198, EVALUATE 1.00e-94
; OTHER INFORMATION: EST_HUMAN HIT: AW206860.1, EVALUATE 2.00e-63

```
; OTHER INFORMATION: SWISSPROT HIT: P20813, EVALUE 2.00e-04
US-10-029-386-25513
Query Match      68.7%; Score 20.6; DB 15; Length 176;
Best Local Similarity 85.2%; Pred. No. 40;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 105 TGGCGCTGAACCCAGGAGTAGCAAGAGT 131

RESULT 3
US-10-029-386-11813
; Sequence 11813, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: ABOMICA-X-2 EXPRESSION ANALYSIS TWO
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11813
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR19.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: 9114572138, EVALUE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: AW206860.1, EVALUE 2.00e-69
; OTHER INFORMATION: SWISSPROT HIT: P20813, EVALUE 2.00e-04
US-10-029-386-11813
Query Match      68.7%; Score 20.6; DB 15; Length 501;
Best Local Similarity 85.2%; Pred. No. 47;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 237 TGGCGCTGAACCCAGGAGTAGCAAGAGT 263

RESULT 4
US-09-954-456-218/c
; Sequence 218, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
```

```
; OTHER INFORMATION: SWISSPROT HIT: P20813, EVALUE 2.00e-04
US-10-029-386-25513
Query Match      68.7%; Score 20.6; DB 15; Length 176;
Best Local Similarity 85.2%; Pred. No. 40;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 105 TGGCGCTGAACCCAGGAGTAGCAAGAGT 131

RESULT 3
US-10-029-386-11813
; Sequence 11813, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: ABOMICA-X-2 EXPRESSION ANALYSIS TWO
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11813
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR19.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: 9114572138, EVALUE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: AW206860.1, EVALUE 2.00e-69
; OTHER INFORMATION: SWISSPROT HIT: P20813, EVALUE 2.00e-04
US-10-029-386-11813
Query Match      68.7%; Score 20.6; DB 15; Length 501;
Best Local Similarity 85.2%; Pred. No. 47;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 237 TGGCGCTGAACCCAGGAGTAGCAAGAGT 263

RESULT 4
US-09-954-456-218/c
; Sequence 218, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
```



```
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 400
; LENGTH: 3045
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-400

Query Match      68.7%; Score 20.6; DB 13; Length 3045;
Best Local Similarity 85.2%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
    |||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAGAGT 48

RESULT 7
US-10-172-118-400/c
; Sequence 400, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van 't Veer, Marc
; APPLICANT: Bernards, Rene
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 400
; LENGTH: 3045
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: M29874
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-400

Query Match      68.7%; Score 20.6; DB 13; Length 3045;
Best Local Similarity 85.2%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
    |||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAGAGT 48

RESULT 8
US-10-342-887-525/c
; Sequence 525, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
```

```
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 525
; LENGTH: 3050
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-525

Query Match      68.7%; Score 20.6; DB 13; Length 3050;
Best Local Similarity 85.2%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
    |||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAGAGT 48

RESULT 9
US-10-172-118-525/c
; Sequence 525, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 525
; LENGTH: 3050
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_000767
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-525

Query Match      68.7%; Score 20.6; DB 13; Length 3050;
Best Local Similarity 85.2%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
    |||||
Db 74 TGGCGCTGAACCCAGGAGTAGCAGAGT 48

RESULT 10
US-10-268-822-12/c
; Sequence 12, Application US/10268822
; Publication No. US20030150004A1
```

GENERAL INFORMATION:
; APPLICANT: Moore, David
; APPLICANT: Wei, Ping
; APPLICANT: Chua, Steven
; TITLE OF INVENTION: Screening Systems and Methods for Identifying Modulators of Xenob
; FILE REFERENCE: P02729US2
; CURRENT APPLICATION NUMBER: US/10/268,822
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US 01/29672
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 10/219,590
; PRIOR FILING DATE: 2002-08-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 42547
; TYPE: DNA
; ORGANISM: Human
US-10-268-822-12

Query Match 68.7%; Score 20.6; DB 15; Length 42547;
Best Local Similarity 85.2%; Pred. No. 89;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGACCCAGGAGTAGCACCACAT 27
Db 4183 TGGCGCTGAACCCAGGAGTAGCAGAGT 4157

RESULT 11

US-10-027-632-22664/c
; Sequence 22664, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22664
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-22664

Query Match 65.3%; Score 19.6; DB 13; Length 629;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GCAGAACCCAGGAGTAGCACCACATGAG 30
Db 224 GCAGAACCCAGGAGTAGCACCAGGAG 199

RESULT 12

US-10-027-632-22664/c
; Sequence 22664, Application US/10027632
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 99934
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)....(99934)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-73

Query Match 65.3%; Score 19.6; DB 16; Length 99934;
Best Local Similarity 84.6%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCAC 26
Db 8930 TGGCGCAGGAGGAGTGCTCCAA 8905

US-10-027-632-22664/c
; Sequence 22664, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22664
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-22664

Query Match 65.3%; Score 19.6; DB 16; Length 629;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GCAGAACCCAGGAGTAGCACCACATGAG 30
Db 224 GCAGAACCCAGGAGTAGCACCAGGAG 199

RESULT 13

US-10-085-117-73/c
; Sequence 73, Application US/10085117
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 99934
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)....(99934)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-73

Query Match 65.3%; Score 19.6; DB 16; Length 99934;
Best Local Similarity 84.6%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCAC 26
Db 8930 TGGCGCAGGAGGAGTGCTCCAA 8905

RESULT 14

US-10-062-674-960/c
; Sequence 960, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
; FILE REFERENCE: PA-0026-1 CIP
; CURRENT APPLICATION NUMBER: US/10/062,674
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: US 09/625,102
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 2217
; SOFTWARE: PERL Program
; SEQ ID NO 960
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US20040005559A1 g4272049
US-10-062-674-960

Query Match 64.7%; Score 19.4; DB 16; Length 360;
Best Local Similarity 79.3%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 30
|||||
DB 332 GGCGCAGAACCCAGGAGTAGCACCACCAAGAG 304
|||||

RESULT 15

US-09-864-761-7559
; Sequence 7559, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7559
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019276.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
US-09-864-761-7559

Query Match 64.7%; Score 19.4; DB 9; Length 537;
Best Local Similarity 79.3%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 30
|||||
DB 49 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 77
|||||

RESULT 16

US-10-371-264-15
; Sequence 15, Application US/10371264
; Publication No. US20030232061A1
; GENERAL INFORMATION:
; APPLICANT: Haller, Aurelia
; APPLICANT: Tang, Roderick
; TITLE OF INVENTION: RECOMBINANT PARAINFLUENZA VIRUS
; TITLE OF INVENTION: EXPRESSION SYSTEMS AND VACCINES
; TITLE OF INVENTION: COMPRISING HETEROLOGOUS ANTIGENS
; TITLE OF INVENTION: DERIVED FROM METAPNEUMOVIRUS
; FILE REFERENCE: 7682-067-999
; CURRENT APPLICATION NUMBER: US/10/371,264
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 60/358,934
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 327
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: rhinotracheitis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)...(1260)
; OTHER INFORMATION: Turkey rhinotracheitis virus (strain 6574)
; OTHER INFORMATION: attachment protein (G), complete cds
US-10-371-264-15

Query Match 64.7%; Score 19.4; DB 16; Length 1260;
Best Local Similarity 79.3%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACCAATGA 29
|||||
DB 1189 TGGCGCAGAACCCAGGAGTAGCACCACCAATGTCGA 1217
|||||

US-10-628-088-15

Query Match 64.7%; Score 19.4; DB 17; Length 1260;
Best Local Similarity 79.3%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGACCAAGGAGTAGCACCATTGA 29
DB 1189 TGGCGAGACCAAGGAGTAGCATTCTGTGA 1217

RESULT 21

US-10-369-493-25221/c
; Sequence 25221, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 25221
; LENGTH: 1779
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-25221

Query Match 64.7%; Score 19.4; DB 16; Length 1779;
Best Local Similarity 79.3%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGACCAAGGAGTAGCACCATTGA 29
DB 276 TGGCGAGACCAAGGAGTAGCACCATTGA 248

RESULT 22

US-10-425-114-15923
; Sequence 15923, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15923
; LENGTH: 1342
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-109-H1_FLI
US-10-425-114-15923

Query Match 64.0%; Score 19.2; DB 13; Length 1342;
Best Local Similarity 87.5%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCGAGTAGCACCA 25
DB 45 GGCGCAGAACCGAGTAGCACCA 68

RESULT 23

US-10-369-493-46459/c
; Sequence 46459, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46459
; LENGTH: 1950
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-46459

Query Match 64.0%; Score 19.2; DB 16; Length 1950;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCGAGTAGCACCA 25
DB 1913 GGCGCAGAACCGAGTAGCACCA 1890

RESULT 24

US-10-158-844-36
; Sequence 36, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:

us-10-624-714-18.rnpb

Mon Jun 21 09:02:14 2004

Query Match 63.3%; Score 19; DB 13; Length 21706;
Best Local Similarity 81.5%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
SEQUENCE DESCRIPTION: SEQ ID NO: 36;
US-10-158-844-36

Query Match 64.0%; Score 19.2; DB 13; Length 21706;
Best Local Similarity 87.5%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGACAGGAGTAGCACCATTG 28
DB 6961 GAAGAACAGAGTAGCTCCATTG 6984

RESULT 25
US-09-918-995-5024/c
Sequence 5024, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5024
LENGTH: 393
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-5024

Query Match 63.3%; Score 19; DB 10; Length 393;
Best Local Similarity 81.5%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGCGCCAGAACCCAGGAGTAGCACCATTG 27
DB 382 TGCGCCAGAACCCAGGAGTAGCTCCGAT 356

RESULT 26
US-09-563-817-317
Sequence 317, Application US/09563817
Patent No. US20020095031A1
GENERAL INFORMATION:
APPLICANT: Nehls, Michael C.
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
TITLE OF INVENTION: No. US20020095031A1 Human Polynucleotides and the
FILE REFERENCE: LEX-0021-USA
CURRENT APPLICATION NUMBER: US/09/563,817
CURRENT FILING DATE: 2001-06-11
PRIOR APPLICATION NUMBER: US 60/132,343
PRIOR FILING DATE: 1999-05-04
NUMBER OF SEQ ID NOS: 1008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 317
LENGTH: 637
TYPE: DNA
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(637)
OTHER INFORMATION: n = A,T,C or G
US-09-563-817-317

Query Match 63.3%; Score 19; DB 9; Length 637;
Best Local Similarity 81.5%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
SEQUENCE DESCRIPTION: SEQ ID NO: 27;
US-10-092-900A-119/c

RESULT 27
US-10-092-900A-119/c
Sequence 119, Application US/10092900A
Publication No. US20040043382A1
GENERAL INFORMATION:
APPLICANT: Padigaru, Muralidhara
APPLICANT: Spytek, Kimberly A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Taupier Jr., Raymond J.
APPLICANT: Pena, Carol E.A.
APPLICANT: Li, Li
APPLICANT: Zerhusen, Bryan D.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Ji, Weizhen
APPLICANT: Gorman, Linda
APPLICANT: Miller, Charles E.
APPLICANT: Kekuda, Ramesh
APPLICANT: Patturajan, Meera
APPLICANT: Gangolli, Esha A.
APPLICANT: Vernet, Corine A.M.
APPLICANT: Guo, Xiaojia Sashu
APPLICANT: Tchernev, Velizar T.
APPLICANT: Fernandes, Elma R.
APPLICANT: Casman, Stacie J.
APPLICANT: Malyankar, Uriel M.
APPLICANT: Gerlach, Valerie
APPLICANT: Liu, Yi
APPLICANT: Anderson, David W.
APPLICANT: Spaderna, Steven K.
APPLICANT: Catterton, Elina
APPLICANT: Leite, Mario W.
APPLICANT: Zhong, Haihong
APPLICANT: Alsobrook, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: NO. US20040043382A1el Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-290C
CURRENT APPLICATION NUMBER: US/10/092,900A
CURRENT FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: USSN 60/274,322
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: USSN 60/283,675
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: USSN 60/338,092
PRIOR FILING DATE: 2001-12-03
PRIOR APPLICATION NUMBER: USSN 60/274,281
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: USSN 60/274,191
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: USSN 60/325,681
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: USSN 60/304,354
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: USSN 60/279,995
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: USSN 60/294,899
PRIOR FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: USSN 60/287,424
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 768
SEQ ID NO 119
LENGTH: 843

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (132)..(771)
US-10-092-900A-119

Query Match      63.3%; Score 19; DB 13; Length 843;
Best Local Similarity 81.5%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGCGCGAGAACCCAGGAGTAGCACCACAT 27
Db 490 TGCGCGAGAACCCAGGAGTAGCACCACAT 464

RESULT 28
US-10-282-174-467/c
; Sequence 467, Application US/10282174
; Publication No. US20030224380A1
; GENERAL INFORMATION:
; APPLICANT: Becker, Kenneth David
; APPLICANT: Velicelebi, Gonul
; APPLICANT: Elliott, Kathryn J.
; APPLICANT: Wang, Xin
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Bertram, Lars
; APPLICANT: Saunders, Aleister J.
; APPLICANT: Mullin, Kristina M.
; APPLICANT: Sampson, Andrew Johnson
; APPLICANT: Blacker, Deborah Lynne
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ON CHROMOSOME 10
; TITLE OF INVENTION: ASSOCIATED WITH ALZHEIMER'S DISEASE AND OTHER
; FILE REFERENCE: 37481-3308
; CURRENT APPLICATION NUMBER: US 60/339,525
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 60/338,010
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/336,929
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/338,363
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/337,052
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/368,919
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 564
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 40178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1197,1841,1852,2075,6063,6173,6194,7219,7820,8242,10114,
; LOCATION: 10606,10688,10729,11559,12031,14437,14729,21145, 21329, 21404,
; OTHER INFORMATION: N is any
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1307,1308,1309
; OTHER INFORMATION: deletion: ATC
; FEATURE:
; NAME/KEY: allele
; LOCATION: (28453)...(28465)
; OTHER INFORMATION: deletion: TCCGCGAGAGGCG
; NAME/KEY: allele
; LOCATION: 21429,22246,22354,23621,23802,25283,25969,29904,39834,40018
; OTHER INFORMATION: N is any
US-10-282-174-468

Query Match      63.3%; Score 19; DB 13; Length 40178;
Best Local Similarity 81.5%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACATG 28
Db 10145 GGCGCAGAACCCAGGAGTAGCACCACAGG 10119

RESULT 29
US-10-282-174-468/c
; Sequence 468, Application US/10282174
; Publication No. US20030224380A1
; GENERAL INFORMATION:
; APPLICANT: Becker, Kenneth David
; APPLICANT: Velicelebi, Gonul
```

```
; APPLICANT: Elliott, Kathryn J.
; APPLICANT: Wang, Xin
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Bertram, Lars
; APPLICANT: Saunders, Aleister J.
; APPLICANT: Mullin, Kristina M.
; APPLICANT: Sampson, Andrew Johnson
; APPLICANT: Blacker, Deborah Lynne
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ON CHROMOSOME 10
; TITLE OF INVENTION: ASSOCIATED WITH ALZHEIMER'S DISEASE AND OTHER
; FILE REFERENCE: 37481-3308
; CURRENT APPLICATION NUMBER: US/10/282,174
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/339,525
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 60/338,010
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/336,929
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/338,363
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/337,052
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/368,919
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 564
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 468
; LENGTH: 40178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1197,1841,1852,2075,6063,6173,6194,7219,7820,8242,10114,
; LOCATION: 10606,10688,10729,11559,12031,14437,14729,21145, 21329, 21404,
; OTHER INFORMATION: N is any
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1307,1308,1309
; OTHER INFORMATION: deletion: ATC
; FEATURE:
; NAME/KEY: allele
; LOCATION: (28453)...(28465)
; OTHER INFORMATION: deletion: TCCGCGAGAGGCG
; NAME/KEY: allele
; LOCATION: 21429,22246,22354,23621,23802,25283,25969,29904,39834,40018
; OTHER INFORMATION: N is any
US-10-282-174-468

Query Match      63.3%; Score 19; DB 13; Length 40178;
Best Local Similarity 81.5%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACATG 28
Db 10145 GGCGCAGAACCCAGGAGTAGCACCACAGG 10119

RESULT 30
US-10-398-221-8/c
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
```

us-10-624-714-18.rnpb

Mon Jun 21 09:02:14 2004

```

; PRIOR APPLICATION NUMBER: PR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 495269
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8

Query Match      63.3%; Score 19; DB 16; Length 495269;
Best Local Similarity 81.5%; Pred. No. 6e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      3 GCGCAGAACCGAGGTAGCACCACCAATGA 29
        |||||
Db      439229 GCGCAGAACCGAGGTAGCACCACCACTGA 439203

Search completed: June 20, 2004, 17:46:47
Job time : 178.683 secs

```